Utrecht University Library
Partner in Science
Universiteit Utrecht

Tijdschrift voor onderwijsresearch 1995

https://hdl.handle.net/1874/208873

TIJDSCHRIFT VOOR ONDERWIJS RESEARCH

ISSN 0166-591 X

Jaargang 20
1995

Swets & Zeitlinger B.V.
Lisse

Redactie:

Klaas J. Westerhof (secretaris)
Christiaan Hamaker
Bernadette van Hout Wolters
Wim J. van der Linden
Joost Lowyck
G. Wim Meijnen
Johan M.M. van der Sanden
Jaap Scheerens
Marinus J.M. Voeten

Universiteit Utrecht
BIBLIOTHEEK CENTRUM UITHOF

SW

VERENIGING VOOR ONDERWIJSRESEARCH

RIJKSUNIVERSITEIT TE UTRECHT

2652 329 4

TIJDSCHRIFT VOOR ONDERWIJSRESEARCH

Een drie-maandelijks tijdschrift voor onderzoekers van onderwijs, gewijd aan theoretische, methodologische en professionele problemen in de onderwijsresearch.

Redactie:
Christiaan Hamaker
Bernadette van Hout Wolters
Wim J. van der Linden
Joost Lowyck
G. Wim Meijnen
Johan M.M. van der Sanden
Jaap Scheerens
Marinus J.M. Voeten
Klaas J. Westerhof (secretaris)

Redactiesecretariaat:
Klaas J. Westerhof
Vakgroep Pedagogiek en Onderwijskunde
Rijksuniversiteit Groningen
Grote Rozenstraat 38
9712 TJ Groningen
tel.: 050-3636492/3636500
fax: 050-3636521
e-mail: K.J. Westerhof@ppsw.rug.nl

Manuscripten zende men in drievoud, getypt en persklaar aan de redactiesecretaris.
Manuscripten dienen te worden voorbereid overeenkomstig de Aanwijzingen voor auteurs.

Abonnementsprijs: Instituten f. 192,50
Particulieren f. 125,00
Studenten f. 102,50
De prijzen zijn inclusief porto, maar exclusief btw.

Men kan zich abonneren bij de uitgever: Swets & Zeitlinger, Heereweg 347, 2161 CA Lisse.
Voor leden van de Vereniging voor Onderwijsresearch (VOR) is het abonnementsgeld reeds inbegrepen in de contributie. Leden van de VOR hoeven zich dan ook *niet* afzonderlijk te abonneren bij de uitgever.

Losse nummers zijn (à f. 50,00) verkrijgbaar bij de uitgever.

Advertenties dienen te worden gezonden aan het adres van de uitgever. Tarieven: 1/1 f. 450,00; 1/2 f. 260,00; 1/4 f. 190,00; exclusief zetkosten.

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotocopie, microfilm of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

Het *Tijdschrift voor Onderwijsresearch* wordt uitgegeven door de Vereniging voor Onderwijsresearch.

Vereniging voor Onderwijsresearch
Secretariaat:
Dr. M. Mulder
Faculteit Toegepaste Onderwijskunde
Universiteit Twente
Postbus 217
7500 AE Enschede
tel.: 053-4893652
fax: 053-4329136
Mulder@edte.utwente.nl

Abonnementenadministratie:
Swets & Zeitlinger
Heereweg 347
2161 CA Lisse
tel. 0252-435111
fax: 0252-415888
giro 13984

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, pp. 1-2

Van de redactie

November 1975 verscheen het eerste nummer van Tijdschrift voor Onderwijsresearch. In datzelfde jaar was ook de Vereniging voor Onderwijsresearch opgericht. Een jaar eerder waren in Amsterdam, met groot succes, de eerste Onderwijsresearchdagen gehouden. Dit waren drie belangrijke stappen in de ontwikkeling van het onderwijsonderzoek in het Nederlandse taalgebied. Vrij snel zijn deze drie initiatieven in elkaar geschoven. Eind zestiger, begin zeventiger jaren was het empirische omderzoek van het onderwijs aan de Nederlandse en Vlaamse universiteiten op gang gekomen. Het waren voornamelijk toen nog jonge en beginnende onderzoekers die – ondersteund door enkele gerenommeerde wetenschappers – via vereniging, congres en tijdschrift een wetenschappelijk platform voor onderwijsonderzoekers hebben gecreëerd. Deze initiatieven zijn manifestaties van (streven naar) een professionele identiteit van onderwijsonderzoekers, ongeacht disciplinaire of interdisciplinaire oriëntatie. Binnen dit geheel heeft het tijdschrift een belangrijke rol vervuld en nog steeds te vervullen.

Tijdschrift voor Onderwijsresearch heeft zich van meet af aan gepresenteerd als een forum voor onderzoek van het onderwijs, waarbij met name veel aandacht werd gegeven aan methodologische en onderzoekstechnische aspecten. Sinds de start in 1975 heeft het tijdschrift zich ontwikkeld tot een publikatiemedium voor onderwijsonderzoek in het algemeen, waarbij een grote variëteit aan inhoudelijke thematiek aan bod kan komen. In de laatste jaren is daarbij een inhoudelijke concentratie tot stand gekomen in de vorm van themanummers.

Sinds 1975 is er veel veranderd in het land van de onderwijsonderzoekers. Naast inhoudelijke ontwikkelingen in het onderzoek en in de wijzen van onderzoeken, waren er veranderingen in onderwijsbeleid en onderwijspraktijk en veranderingen in de instituten en de samenwerkingsverbanden waarbinnen onderzoekers functioneren. Wat het laatste betreft, is er duidelijk sprake van een trend naar internationalisering, die mede tot uiting komt in een streven van onderwijsonderzoekers van verschillende landen om zich te verenigen in de vorm van internationale verenigingen, congressen en tijdschriften. Toen ik het eerste nummer van het Tijdschrift voor Onderwijsresearch weer eens ter hand nam, zag ik dat de eerste helft Engelstalig en de tweede helft Nederlandstalig was. Dat heeft zich niet over alle nummers voortgezet, maar ook nu nog kan men Engelstalige naast Nederlandstalige artikelen en themanummers aantreffen. In vergelijking met de situatie in 1975 zijn Nederlandse en Vlaamse onderwijsonderzoekers meer gaan publiceren in internationale tijdschriften. Dat is ongetwijfeld een goede ontwikkeling, maar het neemt niet weg dat er nog steeds een eigen plaats en taak is voor een publikatiemedium dat door de Nederlandstalige onderwijsresearchwereld wordt gedragen.

Deze korte terugblik op twee decennia onderwijsresearch is vooral ingegeven door het feit dat per 1 januari 1995 een van de belangrijkste hoofdrolspelers afscheid heeft genomen van de redactie. Het is moeilijk om hem uit TOR weg te denken, maar Bert Creemers heeft de redactie verlaten. Zijn vele activiteiten ten behoeve van de onderwijsresearch, waaronder het redigeren van internationale tijdschriften, hebben hem doen besluiten zijn werk voor TOR nu te beëindigen. In de totstandkoming en ontwikkeling van het tijdschrift heeft Bert Creemers een zeer grote rol gespeeld. Hij behoorde tot de oprichters, hij was de initiator. Vanaf het begin is hij lid geweest van de redactie, een aantal jaren als voorzitter. Tien jaar lang, van 1984 tot en met 1994, heeft Bert het werk van redactiesecretaris gedaan. Als redactie danken wij hem voor zijn grote inzet en enthousiasme en voor de prettige wijze waarop wij met hem hebben kunnen samenwerken. De onderwijsonderzoekers in Nederland en Vlaanderen zijn hem veel dank verschuldigd.

Het redactiesecretariaat blijft in Groningen. Bert wordt als redactiesecretaris opgevolgd door Klaas Westerhof. Hij mag beginnen met een ruime erfenis van themanummers en manuscripten die nog onder verantwoordelijkheid van Bert Creemers tot stand zijn gekomen.

Met Bert Creemers heeft ook Annelies Weersing de redactie verlaten. Vanaf 1984 heeft zij Bert bijgestaan in zijn taken als redactiesecretaris en verzorgde zij de notulen van de redactievergaderingen. De redactie wil graag haar grote waardering uitspreken voor de wijze waarop

Annelies in de afgelopen tien jaar het secretariaat van de redactie heeft beheerd. Het was heel plezierig met haar samen te werken. Wij wensen haar veel succes met de organisatie van de ORD'95 in Groningen. Verder heeft de redactie per 1 januari 1995 ook afscheid genomen van Leo van der Kamp als redactielid. Zijn fakkel is overgenomen door Wim van der Linden. Leo maakte sinds november 1984 deel uit van de redactie. Zijn vertrek verhindert niet dat wij er nog steeds van uitgaan ook in de toekomst gebruik te kunnen maken van zijn zeer gewaardeerde bijdragen tot de beoordeling van manuscripten.

Een wetenschappelijk tijdschrift wordt gedragen door auteurs die hun werk aan het tijdschrift aanbieden en door collega's van deze auteurs die het aangeboden werk beoordelen en van commentaar voorzien. In de nu bijna twintigjarige geschiedenis van het tijdschrift heeft de redactie met vele auteurs en vele beoordelaars mogen werken. Wij zijn hen dankbaar voor hun bijdragen. De redactie doet haar best om alle ingezonden manuscripten op een billijke wijze te behandelen en om daarbij zo goed mogelijk gebruik te maken van de in het Nederlandse taalgebied aanwezige expertise. Wij willen u allen aanmoedigen om manuscripten aan ons in te zenden of voorstellen te doen voor themanummers. De redactie wil graag artikelen ontvangen die interessant zijn voor een brede doorsnee van de onderwijsonderzoekers en die een bijdrage leveren aan de voortgang van het onderwijsonderzoek. De redactie vertrouwt erop dat vanuit alle divisies van de Vereniging voor Onderwijsresearch onderzoekers willen meewerken aan de continuering en verdere uitbouw van een onderzoekstraditie die weliswaar nog jong is, maar waar Bert Creemers en wij toch al een beetje trots op beginnen te worden.

Rinus Voeten

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, pp. 3-12

Themanummer: Evaluatie van het basisonderwijs

De evaluatie van het basisonderwijs door de CEB[1]

Frans J.G. Janssens
Inspectie van het Onderwijs[2]

ABSTRACT

In 1992 the Netherlands' Inspectorate of Education appointed a temporary committee (CEB) to evaluate primary education. This evaluation had to be completed within two years. This paper describes the approach and working methods adopted by the CEB and examines certain aspects of the CEB's methodology, such as the development of evaluation criteria, the collection of data and the consultation of experts and others involved in the education sector.
The rationale underlying the choice of quality criteria for the evaluation is a definition of 'good' primary education. The CEB devised frameworks for the evaluation on the basis of this definition: sets of criteria and standards. These frameworks were used to describe and assess the state of education.
The absence of sufficient up-to-date information that was both relevant and methodologically sound, posed a great problem. The CEB regarded it as inadmissible to omit certain important aspects of primary education from the evaluation only on the grounds that too little data were available. It therefore opted to commission research studies. In order to forestall such problems arising during future large-scale evaluations, the CEB recommend that an evaluation plan, including an evaluation framework, be drawn up at a very early stage. On the basis of this plan, a multiyear research and data collection programme should be drawn up.

1. INLEIDING

In 1991 heeft de toenmalige staatssecretaris van O. en W., drs J. Wallage, de inspectie van het onderwijs de opdracht gegeven een tijdelijke commissie te installeren die tot taak kreeg het basisonderwijs te evalueren. Het doel van de evaluatie was, zo formuleerde de staatssecretaris: inzicht krijgen in de feitelijke ontwikkelingen in het basisonderwijs, voor zover deze betrekking hebben op de kerntaken. Aan de commissie is bovendien gevraagd waar mogelijk een oordeel te geven over het niveau en de kwaliteit van het onderwijs en de leerresultaten. In het verlengde hiervan is de commissie verzocht aanbevelingen te doen over de wijze waarop en de richting waarin het basisonderwijs in de toekomst verbeterd kan worden (zie Plan voor de evaluatie van het basisonderwijs: CEB, 1992). Eind 1991 zijn de werkzaamheden van de Commissie Evaluatie Basisonderwijs (CEB) gestart. De afronding van de evaluatie vond, zoals was voorzien, eind 1993 plaats.

Voorliggend artikel beschrijft de aanpak en werkwijze van de CEB. Paragraaf 2 behandelt de aanleiding tot deze evaluatie en de randvoorwaarden waarbinnen de opdracht uitgevoerd moest worden. Paragraaf 3 gaat in op de werkwijze van de CEB en paragraaf 4 op de onafhankelijkheid van deze commissie. In paragraaf 5 wordt een aantal methodologische aspecten van de gekozen evaluatie-aanpak behandeld, zoals de ontwikkeling van evaluatiecriteria, de wijze van gege-

1. Bewerking van een lezing ten behoeve van de Algemene Ledenvergadering van de VOR-divisie Methodologie & Evaluatie, Arnhem, 23 maart 1994.
2. Adres: Kantoor Inspectieleiding, Postbus 134, 3454 ZJ De Meern.

vensverzameling en de raadpleging van experts en bij het onderwijs betrokkenen. In paragraaf 6 worden tot slot enkele kanttekeningen gemaakt bij de uitvoering van de evaluatie, welke van belang kunnen zijn voor eventuele volgende evaluaties.

2. AANLEIDING EN RANDVOORWAARDEN

Bij de inwerkingtreding van de wet op het basisonderwijs (WBO) in 1985 is aangekondigd dat de voortgang van de ontwikkeling van het basisonderwijs nauwlettend zou worden gevolgd. Gedacht werd aan een systematische evaluatie, die om de drie jaar zou plaatsvinden. Het eerste evaluatiemoment lag dan ook in 1988. Deze lijn doortrekkend zou een volgende evaluatie gebaseerd moeten zijn op de stand van zaken in 1991. Hierin is echter vertraging opgetreden. Zowel de kabinetswisseling in 1990 als het grote aantal kritische kanttekeningen bij de eerste evaluatie hebben ertoe geleid, dat de resultaten ervan pas in januari 1991 met de Tweede Kamer konden worden besproken, aan de hand van de nota 'Zo hard gelopen en toch nog zo ver te gaan'.

De belangrijkste kritiek van de zijde van de Tweede Kamer op deze nota betrof de wijze van evalueren. De evaluatie was uitgevoerd door het Ministerie van Onderwijs en Wetenschappen, wat vragen opriep over de objectiviteit ervan. Voor het uitvoeren van een volgende evaluatie werd dan ook door de Kamer gepleit voor een onafhankelijke externe commissie van deskundigen die minder nauw betrokken zijn geweest bij het tot stand komen van de WBO en het basisonderwijs. Bovendien moest een substantiële inbreng van alle onderwijsbetrokkenen gegarandeerd worden. Verder zou een evaluatieplan moeten worden opgesteld.

Een ander punt van kritiek betrof het doel en de inhoud van de eerste evaluatie. Naar het oordeel van velen was die te veelomvattend, te veel gericht op het gevoerde beleid respectievelijk de wet en te weinig op de onderwijspraktijk. Daarom werd gepleit voor een systematische evaluatie van de feitelijke ontwikkelingen in het basisonderwijs, waarbij men zich zou moeten beperken tot enkele kernproblemen.

De suggesties van de Tweede Kamer, eveneens verwoord in de opdrachtformulering van de staatssecretaris (zie daarvoor: Plan voor de evaluatie van het Basisonderwijs; CEB, 1992), zijn richtinggevend geweest bij de opzet van de recente evaluatie. Naast het verzoek internationale expertise in te schakelen en gebruik te maken van over het basisonderwijs beschikbare onderzoeksgegevens, zijn ook de volgende voorwaarden van belang;

1. *De evaluatiecommissie moest onafhankelijk zijn*
   De onafhankelijkheid werd volgens de staatssecretaris gegarandeerd door de instelling van een tijdelijke commissie, onder leiding van de Inspectie van het onderwijs. Bovendien zou deze Commissie Evaluatie Basisonderwijs (CEB) samengesteld moeten zijn uit onafhankelijke deskundigen die geen directe betrokkenheid hebben (gehad) met de beleidsontwikkeling en -uitvoering op het terrein van het basisonderwijs. In de CEB hebben, onder voorzitterschap van de Inspecteur-generaal van het onderwijs, A.J.C. Stekelenburg, zitting genomen de hoogleraren B.P.M. Creemers, mw L. Eldering, A. van der Leij, W.J. van der Linden en J. Scheerens.

2. *Het onderwijsveld moest bij de evaluatie worden betrokken*
   Om een substantiële inbreng te garanderen van betrokkenen bij het basisonderwijs heeft de CEB diverse groeperingen geraadpleegd over zowel de evaluatievragen als de gehanteerde criteria en de bereikte resultaten (zie ook par. 5.4).
   De betrokkenheid van de onderwijsvakorganisaties, de besturenorganisaties en de ouderorganisaties is belichaamd in een begeleidingscommissie van de Centrale Commissie voor Onderwijsoverleg (CCOO), die het hele traject heeft gevolgd, van planopstelling tot rapportage.

3. *De evaluatie moest betrekking hebben op de feitelijke ontwikkelingen op het gebied van de kernproblemen in het basisonderwijs*
De inhoudelijke suggesties van de Tweede Kamer naar aanleiding van de eerste evaluatie van het basisonderwijs bevatten twee elementen: 'feitelijke ontwikkelingen' en 'kernproblemen van het basisonderwijs'. Dit is gehonoreerd door deze evaluatie op te zetten als een waarderende beschrijving van de toestand van het basisonderwijs toegespitst op de door de staatssecretaris geformuleerde centrale vragen. Deze luiden: Hoe staat het ervoor met betrekking tot de volgende aandachtsgebieden:
1. de inhoud en opbrengsten van het onderwijs?
2. onderwijs-op-maat?
3. het onderwijs aan jonge kinderen?
4. het onderwijs gericht op een multiculturele samenleving?

Deze aandachtsgebieden zijn in vier afzonderlijke deelrapporten uitgewerkt. Daarnaast is nog een eindrapport geproduceerd 'Zicht op kwaliteit', waarin de conclusies uit de deelrapporten opnieuw gepresenteerd zijn en in een onderling verband zijn gebracht. Verder zijn er aanbevelingen gedaan voor verbetering van het basisonderwijs.

3. WERKWIJZE

In de CEB is materiedeskundigheid aangezocht op het gebied van de door de staatssecretaris aangereikte centrale evaluatievragen, alsmede expertise op het gebied van onderwijsevaluatie, evaluatie-onderzoek in het algemeen en internationaal vergelijkend onderzoek. Doordat de CEB onder voorzitterschap stond van de Inspecteur-generaal, kon bovendien specifieke (veld)kennis en ervaring van de Inspectie van het onderwijs worden ingebracht. Het gaat hierbij om zowel de inhoudelijke materiedeskundigheid ten aanzien van de vier aandachtsgebieden als om de kennis van het functioneren van scholen.

De CEB als geheel is verantwoordelijk voor de opzet en uitvoering van de evaluatie, alsmede voor de daaruit voortvloeiende rapporten. Daarnaast zijn de onderscheiden leden van de CEB, op basis van hun specifieke expertise, verantwoordelijk voor het totstandbrengen van een of meer deelrapporten.

Ten behoeve van de inhoudelijke, organisatorische en administratieve ondersteuning van de CEB is door de inspectie een secretariaat geïnstalleerd, dat onder andere bestond uit inspecteurs die deskundig zijn op de onderscheiden aandachtsgebieden.

De CEB heeft haar werkzaamheden verricht op basis van een, op het eerder genoemde evaluatieplan gebaseerd, operationeel werkplan. In grote lijnen is dit werkplan voor de onderscheiden aandachtsgebieden op de volgende stappen gebaseerd:

1 Verrichten van literatuuronderzoek met het oog op enerzijds het formuleren van de criteria waarlangs wordt geëvalueerd en anderzijds het opsporen van bruikbare evaluatiegegevens;
2. Formuleren van evaluatiecriteria (waarderingskader);
3. – Verzamelen van bestaande gegevens en beoordelen op kwaliteit en bruikbaarheid;
   – Uit laten voeren van onderzoek om nieuwe/aanvullende gegevens te verzamelen;
4. Opstellen van een normering om te kunnen vaststellen in welke mate aan de evaluatiecriteria is voldaan;
5. Opstellen van de rapportage.

4. ONAFHANKELIJKHEID VAN DE CEB

Op verzoek van de Tweede Kamer en de verantwoordelijke bewindslieden is de Evaluatie van het basisonderwijs uitgevoerd door een onafhankelijke commissie van de Inspectie van het Onderwijs.

De onafhankelijkheid van de CEB komt tot uitdrukking in het feit, dat:

1. de evaluatiecriteria zijn gebaseerd op opvattingen over de intrinsieke kwaliteit van het basisonderwijs;
2. de oordelen van de CEB niet ontleend zijn aan opinies van bij het basisonderwijs betrokkenen of die van beleidsverantwoordelijken, maar voortvloeien uit het perspectief van waaruit de Inspectie evalueert, nl. het algemeen belang.

Dit stemt overeen met de opvattingen van de Commissie Programma-evaluatie, zoals verwoord in het slotdocument (Laros, 1992, p. 105): 'Onafhankelijke evaluatie houdt in dat getracht wordt criteria af te leiden uit de intrinsieke doelstellingen van een program en uit het mandaat van de beleidsmaker, en dat criteria niet op gezag worden aanvaard. (....) Het onafhankelijkheidscriterium houdt in dat een weliswaar feilbare, maar getrouwe poging wordt ondernomen te oordelen vanuit het gezichtspunt van de staatsburger (niet te verwarren met de consument van het program) die de beleidsmaker een bepaald mandaat (bijvoorbeeld de zorg voor de kwaliteit van het onderwijs) heeft verstrekt'.

De Inspectie van het onderwijs is een onder de eindverantwoordelijkheid van de minister opererend orgaan, onder andere belast met de waarneming en de beoordeling van de toestand van het onderwijs. De inspectie kent een grote onafhankelijkheid in relatie tot de Minister van Onderwijs en Wetenschappen en het departement, dat wil zeggen dat de inspectie zelfstandig en onder eigen verantwoordelijkheid werkt en rapporteert. De inspectie is in staat de toestand van het onderwijs te evalueren, omdat zij niet deelneemt aan en mede-verantwoordelijk is voor de formulering van het overheidsbeleid. Immers, wanneer dat wel het geval zou zijn, is er sprake van twee onverenigbare activiteiten: beleid maken en de effecten van datzelfde beleid evalueren.

De rapportages die de inspectie gevraagd en ongevraagd uitbrengt zijn openbaar. Bovendien is de inspectie verantwoordelijk voor het jaarlijks uit te brengen onderwijsverslag. Ook daarin komt de onafhankelijkheid van de inspectie tot uitdrukking.

5. EVALUATIE-AANPAK

5.1. Begrippenkader

Wetenschappelijk onderzoek en evaluatie hebben veel gemeenschappelijk, met name de systematische gegevensverzameling, maar verschillen op één belangrijk punt van elkaar: onderzoek is vooral gericht op uitbreiding van generaliseerbare kennis met het oog op theorievorming (zie bijv. Wolf, 1990, p. 10-11; Worthen & Sanders, 1987, p. 26 e.v). Bij evaluatie gaat het om een oordeel over de waarde van 'iets'. Daarvoor zijn enerzijds criteria nodig, waarlangs kan worden vastgesteld of het object van evaluatie is gerealiseerd. Anderzijds zijn er normen nodig om vast te kunnen stellen of en in welke mate aan die criteria wordt voldaan. Criteria en normen zijn een eerste noodzakelijke voorwaarde voor een goede evaluatie (De Groot, 1986; Scheerens, 1983; Van den Berg, 1987; Corporaal, Janssens, Sandbergen & Zwarts, 1992; Van den Berg, Hoeben & Van der Werf, 1993).

5.1.1. Kwaliteitscriteria

Een van de belangrijkste onderdelen van een evaluatie is de vaststelling van de criteria waaraan het te evalueren object moet voldoen, wil er sprake zijn van voldoende kwaliteit. Voor het typeren van de kenmerken van en kwalificaties voor te evalueren objecten zijn verschillende begrippen in omloop: criteria, standaarden, maatstaven en normen (zie bijv. Creemers, Hoeben, Peschar & Snippe, 1993; Janssens & Zwarts, 1993). Het onderscheid tussen kenmerken en kwalificaties wordt vaak impliciet gelaten. Zo hanteert bijvoorbeeld het Nederlands Normalisatie-instituut uitsluitend het begrip 'norm' als 'kenmerk inclusief kwalificatie'. Bovendien worden de bovengenoemde termen vaak door elkaar gebruikt. Dezelfde termen worden gebruikt om zowel de lat waarmee gemeten wordt als de onderscheidingen die op die lat worden gemaakt, aan te duiden. Omdat het gebruik van de bedoelde termen, noch taalkundig noch wetenschappe-

lijk gezien, eenduidig is, heeft de CEB twee begrippen gekozen: 'kwaliteitscriterium' om de kritische componenten van het object van kwaliteit te definiëren en 'norm' om de mate waarin de kwaliteitscriteria in de onderwijspraktijk worden gerealiseerd van een kwalificatie te voorzien (Janssens & Zwarts, 1993).

Meestal is het niet mogelijk rechtstreeks de realisatie van een kwaliteitscriterium vast te stellen. Daarvoor heeft de CEB gebruik gemaakt van 'kwaliteitskenmerken' die aangeven welke aspecten essentieel zijn voor de realisering van een kwaliteitscriterium. Deze kwaliteitskenmerken zijn te beschouwen als de factoren die de kwaliteitscriteria operationaliseren. De relatie tussen object van evaluatie, kwaliteitscriteria en -kenmerken, kan worden weergegeven als een hiërarchische conceptuele structuur, die steeds verder wordt uitgewerkt. Een voorbeeld kan dit wellicht verduidelijken: bij de evaluatie van de inhoud en opbrengsten van het basisonderwijs heeft één van de kwaliteitscriteria betrekking op de schoolorganisatie en het pedagogisch klimaat: de schoolorganisatie en het pedagogisch-didactisch klimaat van de school zijn gunstig voor de bevordering van het onderwijsleerproces. Dit kwaliteitscriterium is uiteengelegd in drie kwaliteitskenmerken:

- Er is een duidelijk consistent en coherent schoolbeleid ten aanzien van de pedagogisch-didactische aanpak;
- De school en de schoolomgeving bieden de leerlingen een ordelijk en veilig klimaat;
- Er is een goede relatie tussen de school en de ouders/verzorgers, het voortgezet onderwijs en met externe instanties.

De realisatie van elk van deze kenmerken is vervolgens verder uiteengelegd in subkenmerken en gemeten met behulp van meerdere variabelen.

5.1.2. Normen

Normen geven de beslissingsregels en de kwalificaties voor de mate waarin aan kwaliteitscriteria wordt voldaan. Er zijn verschillende soorten normen, die op verschillende manieren gebruikt kunnen worden. Het gebruik hangt vooral af van de evaluatievraag. Bij internationaal evaluatieonderzoek, zoals bijvoorbeeld uitgevoerd door de IEA, worden landen onderling met elkaar vergeleken. In dat geval worden relatieve normen gehanteerd, waarbij een kwalificatie wordt toegekend aan de positie in een verdeling.

De evaluatie van het basisonderwijs is echter gericht op het beoordelen van de intrinsieke kwaliteit van een complete sector van het onderwijssysteem. In dat geval ligt het hanteren van absolute normen voor de hand. Dit houdt in dat kwalificaties worden toegekend op basis van een inhoudelijke redenering over de mate waarin bepaalde kwaliteitscriteria gerealiseerd zouden moeten zijn. De sterke kant van een dergelijke wijze van normeren is het expliciete karakter van de basis waarop is genormeerd. De zwakke kant van absoluut normeren is dat het tot op zekere hoogte een arbitraire kwestie blijft. Vooral het bepalen van de cesuur is en blijft 'mensenwerk'. Over het algemeen heeft de CEB drie kwalificaties gebruikt: 'onvoldoende' als de kwaliteit onder de maat is en er dus aanleiding is om actie te ondernemen, 'matig' als er twijfel is over de kwaliteit en 'voldoende' als de kwaliteit goed is en er geen reden is om actie te ondernemen. Bij het bepalen van de cesuur (= grens tussen onvoldoende en matig) is door de CEB de volgende werkwijze toegepast. De normen zijn gebaseerd op een beredeneerde keuze uit de (gewogen) geoperationaliseerde kenmerken van een kwaliteitscriterium. Daarbij is rekening gehouden met relevante informatie uit andere bronnen, zoals empirische gegevens, expertise van de inspectie, informatie uit raadplegingen, e.d.

Bij het opstellen van normen is de CEB uitgegaan van de minimale eisen die aan de realisering van de kwaliteitscriteria gesteld kunnen worden, omdat het van belang is in ieder geval te kunnen vaststellen of het onderwijs op een acceptabel niveau wordt gerealiseerd en georganiseerd. Uitgaan van het hoogste bereikbare niveau levert het probleem op dat in de praktijk deze hoge normen in veel gevallen niet (voor alle kwaliteitscriteria) worden bereikt, waardoor de beeldvorming over het onderwijs (onbedoeld) negatief beïnvloed wordt.

5.2. Ontwikkeling van waarderingskaders

De selectie van kwaliteitscriteria is geen eenvoudige zaak, omdat deze geen objectief gegeven zijn, maar beredeneerd gekozen worden. Bij die keuze spelen het perspectief van waaruit geëvalueerd wordt en de belangen die in het spel zijn, een grote rol. Als rationale voor de keuze van kwaliteitscriteria voor de evaluatie van het basisonderwijs gold een omschrijving van 'goed' onderwijs. De CEB meent dat goed basisonderwijs bepaald wordt door (zie CEB, 1992):

1. een leer- en vormingsaanbod
   - dat maatschappelijk relevant is en aangepast is aan de eisen die daaraan thans gesteld moeten worden;
   - dat vervat is in onderwijskundig verantwoord onderwijsleermateriaal;
   - waarvoor voldoende tijd om te leren wordt geboden;
2. hoge opbrengsten die op een efficiënte wijze zijn verkregen;
3. gunstige omstandigheden op het gebied van instructie, schoolorganisatie en pedagogisch klimaat;
4. gunstige omstandigheden op het gebied van
   - onderwijswetgeving, onderwijsbeleid en financiële middelen;
   - materiële voorzieningen;
   - leraarpopulatie.

De mate waarin scholen beoogde opvoedingsdoelen realiseren komt in deze evaluatie niet expliciet aan de orde. Specifieke opvoedingsdoelen behoren immers niet tot het domein van de regelgevende overheid. De formulering van de algemene doelen van het basisonderwijs in de WBO sluiten echter wel opvoedingsdoelen in. Deze algemene doelen maken onderdeel uit van de evaluatie. Datzelfde geldt voor de beschrijving en beoordeling van het pedagogisch klimaat. Ook in het pedagogisch klimaat liggen opvoedingsdoelen besloten.

De CEB heeft waarderingskaders voor de vier aandachtsgebieden ontwikkeld op basis van bovengenoemde omschrijving van goed basisonderwijs. Onder 'waarderingskader' verstaat de CEB een set van kwaliteitscriteria en -kenmerken met bijbehorende normen voor een bepaald aandachtsgebied. Deze kaders zijn gebruikt om voor ieder aandachtsgebied de toestand van het onderwijs te beschrijven en te waarderen.
Uitgangspunt bij het formuleren van kwaliteitscriteria respectievelijk kwaliteitskenmerken en bij het vaststellen van de normen is de intrinsieke kwaliteit van het basisonderwijs: wat wordt het basisonderwijs geacht op te leveren. Vanwege het kaderstellende karakter van de WBO en de verdere beleidsuitwerking daarvan lag het primaat bij het perspectief van de wetgever. Deze wet bevat immers de eisen waaraan basisscholen moeten voldoen, wil er sprake zijn van de realisering van goed onderwijs.

Bij de verdere uitwerking van de waarderingskaders heeft de Commissie zich laten leiden door twee andere perspectieven: dat van de wetenschap en dat van de direct bij het onderwijs betrokkenen. De kwaliteitscriteria zijn mede gebaseerd op de uitkomsten van wetenschappelijk onderzoek naar factoren die wel of juist niet bevorderlijk zijn voor goed onderwijs. En ten slotte zijn ook de opvattingen over goed basisonderwijs van onderwijsgevenden, schoolbesturen, ouders en andere betrokkenen van invloed geweest.
De waarderingskaders zijn tot stand gekomen op basis van een conceptuele analyse van de kwaliteitsbepalende factoren van de onderscheiden aandachtsgebieden. Ten behoeve van de nadere specificering en nuancering van de waarderingskaders is, vaak heen en weer gaand, gebruik gemaakt van wetenschappelijke kennis, wet- en regelgeving, expertise van de inspectie en andere deskundigen. Voorts zijn deze kaders ter validering en legitimering voorgelegd aan verschillende deskundigen en vertegenwoordigers van belangengroeperingen (zie par. 5.4).

5.3. Gegevensverzameling

De Commissie heeft zo veel mogelijk gebruik gemaakt van reeds beschikbare onderzoeksgegevens over de toestand van het Nederlandse basisonderwijs. Zij is daarbij al snel gestuit op het probleem, dat voor een vrij groot aantal onderdelen van de evaluatie onvoldoende recente

onderzoeksresultaten beschikbaar waren. Deze bleken noodzakelijk om tot gefundeerde uitspraken te komen. Aanvullende gegevens waren dus nodig.

5.3.1. Criteria voor selectie van onderzoeksgegevens
Richtinggevend voor het opsporen van bruikbaar bestaand onderzocksmateriaal waren:
1. de bij de opdrachtformulering van de staatssecretaris vermelde centrale vragen bij de vier aandachtsgebieden (zie CEB, 1992);
2. de omschrijving van goed onderwijs, zoals uitgewerkt in het evaluatieplan;
3. de onderscheiden waarderingskaders.

Aan de selectie van het onderzoeksmateriaal lagen de volgende criteria ten grondslag:

1. *Relevantie*
   Het onderzoek moest betrekking hebben op de feitelijke toestand van het onderwijs en niet langer dan 5 à 6 jaar geleden gepubliceerd zijn.
2. *Kwaliteit*
   Het onderzoek moest voldoen aan de gebruikelijke eisen van betrouwbaarheid en validiteit.

5.3.2. Ontbrekende informatie
De CEB constateerde dat op verschillende terreinen onvoldoende bruikbare onderzoeksgegevens beschikbaar waren, waardoor de mogelijkheid tot gefundeerd oordelen over de toestand van het basisonderwijs ernstig werd bedreigd. De belangrijkste hiaten betreffen:

Aanbod
Er was onvoldoende materiaal voorhanden op basis waarvan uitspraken gedaan konden worden over wat leerlingen feitelijk op de basisschool onderwezen werd. Dit gold zowel voor het onderwijsaanbod in de gangbare onderwijsmethoden, als voor het feitelijke realisering van het aanbod in de onderwijspraktijk.

Er waren wel verschillende publikaties beschikbaar waarin, vooral recente, methodes werden beoordeeld, maar deze beoordelingen kenden geen uniforme systematiek. Bovendien stemden de beoordelingscriteria vaak niet overeen met de criteria die in CEB-verband zijn aangelegd om het aanbod te beoordelen. De CEB-criteria hadden betrekking op: de overeenstemming met de kerndoelen, de onderwijskundige kwaliteit en het eigentijdse karakter (zie CEB, 1994).

Ook het bestaande onderzoek naar de implementatie van curricula bleek niet geschikt om het vigerende aanbod in het basisonderwijs te beschrijven en te beoordelen.
Het onderzoek naar het vigerende aanbod dat in PPON-verband is uitgevoerd, bleek niet vrij van methodologische problemen (zie Harskamp & Suhre, 1993).

Onderwijs-op-maat
Over twee belangrijke onderdelen van onderwijs-op-maat bleken eveneens in onvoldoende mate onderzoeksgegevens beschikbaar: differentiatie en zorgverbreding. Het onderzoek naar 'differentiatie' bleek grotendeels betrekking te hebben op de situatie voor de start van het basisonderwijs (zie Stokking, 1991), terwijl er naar essentiële onderdelen van de zorgverbredingsproblematiek praktisch geen grootschalig onderzoek gedaan was.

Pedagogisch klimaat
Ook over dit aspect van het basisonderwijs bleken onvoldoende recente gegevens beschikbaar, zoals over de pedagogische visie en doelstellingen, het pedagogisch handelen en de fysieke omgeving.

Onderwijs aan jonge kinderen
In bovenstaande gevallen ging het om ontbrekend materiaal over belangrijke aspecten van een of meer aandachtsgebieden. Ten aanzien van het aandachtsgebied 'Onderwijs aan jonge kinderen' bleek dat over de gehele linie de nodige empirische gegevens ontbraken. Dit betrof zowel het aanbod, het onderwijsproces als de opbrengsten van het onderwijs aan 4 tot 8 jarigen.

Tabel 1. Aanvullend onderzoek ten behoeve van de evaluatie basisonderwijs.

Onderwerp	Uitvoerende instantie
Literatuuronderzoek	RION via SVO
Methodebeoordeling	RION via SVO
Feitelijk onderwijsaanbod	RION/inspectie
Pedagogisch klimaat	OCTO/inspectie
Onderwijs-op-maat	ISOR via SVO
Onderwijs jonge kinderen	Cito/inspectie
Secundaire analyses	ISOR/inspectie

Procesgegevens
De CEB wenste ook het onderwijsproces en de organisatie van het basisonderwijs te beoordelen. Over belangrijke aspecten daarvan (bijv. het didactisch handelen, het schoolbeleid en de relatie school-omgeving) bleken eveneens belangrijke recente gegevens te ontbreken.

Al vrij snel besloot de CEB naar bovenstaande onderwerpen nieuw of aanvullend onderzoek te laten uitvoeren (zie tabel 1). Een deel van het onderzoek is via de SVO door verschillende onderzoeksinstituten uitgevoerd, al dan niet in samenwerking met de inspectie. Een deel is geheel uitgevoerd door de inspectie zelve.

5.4. Raadplegingen

De CEB heeft om uiteenlopende redenen in verschillende fasen van de evaluatie ruim 260 personen geraadpleegd. Het betrof hier raadplegingen die tot doel hadden:
1. gegevens ten behoeve van de evaluatie te verzamelen;
2. de evaluatie-aanpak te valideren en/of te legitimeren;
3. het verwerven van een draagvlak in de kring van betrokkenen en belanghebbenden.

De inspectie heeft samen met de CEB in twee fasen een groot aantal personen uit scholen voor basisonderwijs, speciaal onderwijs en voortgezet onderwijs geraadpleegd: directieleden, leraren en bestuursleden.

Voorts zijn in overleg met de Vereniging van Onderwijsbegeleidingsdiensten (WPRO) en de directies van de landelijke pedagogische centra vertegenwoordigers aangewezen, die op persoonlijke titel hebben deelgenomen aan raadplegingen. Ditzelfde geldt voor het Instituut voor Toetsontwikkeling (Cito) en het Instituut voor Leerplanontwikkeling (SLO).

De speciale plaats van de ouders is gehonoreerd door aan de ouderorganisaties te vragen personen voor de raadplegingen aan te wijzen.

Verder zijn over specifieke onderwerpen besprekingen gevoerd met vertegenwoordigers van de Landelijke Advies- en Overlegstructuur Minderhedenbeleid (LAO), allochtone ouders, leraren die onderwijs in de eigen taal (en cultuur) verzorgen, inspecteurs, vertegenwoordigers van de vroegere zorgverbredingsprojecten en specifieke deskundigen.

De gevolgde werkwijze, de gehanteerde waarderingskaders en de voorlopige uitkomsten van de evaluatie zijn voorgelegd aan een forum van buitenlandse deskundigen.

De eerste raadplegingen in 1992 gingen over het gekozen evaluatieperspectief. Op grond hiervan is het operationele werkplan van de CEB op een aantal punten gewijzigd. Zo is meer aandacht geschonken aan niet-cognitieve aspecten van het leer- en vormingsaanbod en is het pedagogisch klimaat expliciet opgenomen in de definitie van goed basisonderwijs. Ook de deskundigheid van leraren en de rol van de ouders zijn meer nadrukkelijk bij de evaluatie betrokken. Datzelfde geldt voor de aandacht voor hoogbegaafde leerlingen, de zorgverbredingsprojecten en het ononderbroken ontwikkelingsproces.

De tweede serie raadplegingen, uitgevoerd in 1993, was gericht op de voorlopige uitkomsten van de evaluatie. De algemene tendens van deze raadplegingen was tweeledig. In de eerste plaats werden de conclusies van de Commissie onderschreven; volgens een overgrote meerder-

heid van de betrokkenen geven deze conclusies een juist beeld van de kwaliteit van het basisonderwijs op dit moment. Maar daarnaast werd het als een belangrijke omissie gezien, dat de rapportages geen verklaring gaven van de huidige toestand van het basisonderwijs. Op onderdelen nuanceerden de deelnemers aan de raadplegingen de conclusies van de Commissie. Deze opmerkingen en suggesties hebben de Commissie extra aanleiding gegeven de rapporten aan te vullen met een verklaring van de situatie waarin het basisonderwijs verkeerde.

6. BESLUIT

De beschikbaarheid van voldoende actuele, relevante en methodologisch verantwoorde informatie vormde een groot probleem voor de CEB. De commissie achtte het niet aanvaardbaar om belangrijke aspecten van het basisonderwijs in de evaluatie onbesproken te laten, omdat daarvoor geen (voldoende) gegevens beschikbaar waren. De CEB heeft er toen voor gekozen onderzoek uit te zetten om deze lacunes aan te vullen. Er moest in een kort tijdsbestek (de evaluatie had een looptijd van twee jaar) nog het nodige onderzoek worden uitgezet. Het gevolg hiervan was dat bepaalde onderzoeken onder druk van deze tijdlimiet beperkt van opzet moesten zijn.

Om dergelijke problemen bij toekomstige grootschalige evaluaties te ondervangen, heeft de CEB in het eindrapport een aanbeveling opgenomen waarin wordt gepleit reeds aan het begin van de volgende evaluatieperiode een beknopt evaluatieplan, inclusief een waarderingskader, samen te stellen. Vervolgens zou op basis van dit plan een meerjarige programmering van onderzoek en gegevensverzameling tot stand moeten komen. Concreet betekent dit afstemming van onderzoek, gegevensverzameling en evaluatie door de inspectie en periodieke peilingen van het onderwijsniveau door het Cito. Ook is afstemming gewenst met andere evaluatie-activiteiten, zoals de Landelijke Evaluatie van het Onderwijsvoorrangsbeleid en de landelijke Stuurgroep Evaluatie Weer Samen Naar School.

Naast periodieke peilingen van het onderwijsniveau zou een onderzoeksprogrammering in ieder geval moeten voorzien in een periodieke en systematische verzameling van gegevens over aanbod, context en processen in het basisonderwijs. Verder is systematische vergelijking gewenst van peilingsresultaten, tenminste in EG-verband.

Eind jaren negentig is de evaluatie voorzien van de basisvorming. Teneinde bovenstaande problemen te voorkomen wordt in dat verband thans gewerkt aan de programmering van onderzoek en gegevensverzamelingen. Voorts wordt ernaar gestreefd tijdig het waarderingskader en het evaluatieplan te ontwikkelen.

LITERATUUR

Berg, G. van den (1987). *Effectief evalueren: een empirische studie naar de doelmatigheid van aanwijzingen voor de evaluatiepraktijk.* Lisse: Swets & Zeitlinger.

Berg, G. van den, Hoeben, W.Th.J.G. & Van der Werf, G. (1993). De bijdrage van evaluatie-onderzoek aan de kwaliteitsverbetering van het onderwijs. In B.P.M. Creemers, W.Th.J.G. Hoeben, J.L. Peschar & J. Snippe (1993). *Wat is onderwijsonderzoek waard?* De Lier: Academisch Boekencentrum.

Commissie Evaluatie Basisonderwijs (1992). *Plan voor de evaluatie van het basisonderwijs.* Zoetermeer: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994). *Inhoud en opbrengsten van het basisonderwijs.* De Meern: Inspectie van het Onderwijs.

Corporaal, A.H., Janssens, F.J.G., Sandbergen, S. & Zwarts, M. (1992). *Evaluatie door de inspectie.* Zoetermeer: Inspectie van het Onderwijs

Creemers, B.P.M., Hoeben, W.Th.J.G., Peschar, J.L., & Snippe, J. (1993). *Wat is onderwijsonderzoek waard?* De Lier: Academisch Boekencentrum.

Groot, A.D. de (1986). *Begrip van evalueren.* Den Haag: VUGA.

Harskamp, E. & Suhre, C. (1993). Psychometrische kwaliteiten van meetinstrumenten van het onderwijsaanbod. *Tijdschrift voor Onderwijsresearch,* 17, 355-369.

Inspectie van het Onderwijs (1994). *Technisch rapport van het onderzoek onder 386 basisscholen ten behoeve van de evaluatie van het basisonderwijs.* De Meern; Inspectie van het Onderwijs.

Janssens, F.J.G. & Zwarts, M. (1993). Het gebruik van 'standaarden en normen' binnen de inspectie. *Introspectie,* 14 (1) p. 7-8.

Laros, J.A. (1992). *Eindverslag van de Commissie Programma-evaluatie.* Zoetermeer: Ministerie van Onderwijs en Wetenschappen.

Scheerens, J. (1983). *Evaluatie-onderzoek en beleid: methodologische en organisatorische aspecten.* Den Haag: SVO.

Stokking, K.M. (1991). *Wijkende condities. Differentiatie en zorgverbreding in het basisonderwijs: onderzoek en beleid.* Utrecht: ISOR.

Wolf, R.M. (1990). The nature of educational evaluation. In H.J. Walberg & G.D. Haertel. *The international encyclopedia of educational evaluation.* Oxford: Pergamon Press.

Worthen, B.R. & Sanders, J.R. (1987). *Educational evaluation: alternative approaches and practical guidelines.* New York: Longman.

Manuscript ontvangen 9-8-1994
Definitieve versie ontvangen 31-10-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, pp. 13-27

Robustness of judgments in evaluation research

W.J. van der Linden[1]* and M.A. Zwarts[2]
[1]*University of Twente;* [2]*Inspectorate of Education*

ABSTRACT

The point of view is taken that judgments in evaluative research are ultimately subjective but that good criteria are available to assess their quality. One of these criteria is robustness of the judgments against incompleteness or uncertainty in the data used to describe the educational system. The use of the robustness criterion is demonstrated for the case of a recent evaluation project in which the state of elementary education in The Netherlands was evaluated.

Typically, the first stage of an evaluation project consists of a careful description of the state of an educational object or system. In the next stage, the state of the system is evaluated through a series of evaluative statements or *judgments*. Examples of such judgments are: "The quality of teaching in the system is excellent"; "Too many students in the system do not reach a satisfactory level of proficiency in physics"; and "School management is poor". If the goal of the evaluation project is to serve a reorientation of a policy with respect to the system, the judgments usually result in a series of recommendations to improve the functioning of the system.

For the descriptive stage, the standard methodology of empirical research in the social sciences is available. This methodology includes the use of such methods as survey and observation as well as various techniques of (multivariate) descriptive statistical analysis to summarize the results. Though descriptive statements can be founded on a rigorous methodology, judgments seem to lack this support. The main reason is the use of such qualifications as "excellent", "not satisfactory", and "poor" in the examples above. The choice of such qualifications, as well as their definitions, is a subjective matter. However, subjectivity is not necessarily erratic, and criteria for good qualifications do exist. Judgment does not imply lack of rationality.

One criterion for the quality of judgments is *consistency*. For example, suppose that empirical research has shown time and again that certain instructional measures lead to an increase in the achievements of the students in a given domain, and that a system to be evaluated scores high on the use of these measures. Then, ignoring the role of costs as well as the possibility of interaction between factors in the system, it seems inconsistent to make judgments that provide the former finding with a negative and the latter with a positive qualification. Such evaluations are inconsistent in the sense that they imply a world that can never exist. It should be noted that in this example empirical research was used to show that a set of qualifications is inconsistent. Empirical research can only provide the evaluator with objective information about what worlds are possible and what not. It remains a subjective choice to evaluate one possible world over the other.

Another obvious criterion is *explicitness*. The criterion of explicitness includes the requirement that all judgments be based on explicit definitions of the qualifications and procedures used in the evaluation. If this requirement is not met, the evaluator can never communicate his evaluations to others in a meaningful way. Also, it will never be possible to test these evaluations for consistency in the sense defined above.

Both authors are equally reponsible for the contents of this article; the order of their names is alphabetical. The empirical results in this article were presented earlier at a Vereniging voor Onderwijsresearch, Divisie Methodologie en Evaluatie, symposium, Arnhem, March 23, 1994.

* Adres: Universiteit Twente, Faculteit Toegepaste Onderwijskunde, Postbus 217, 7500 AE Enschede.

It is not the purpose of this paper to give an extensive overview of criteria for the use of qualifications in evaluation research (for a more complete review, see van der Linden, in press). Rather, the emphasis is on one criterion of a more technical nature than the previous examples. The criterion is necessary because judgments may have to be based on a description of the state of the system which is incomplete, uncertain, or erroneous due to the quality of the data. An example is an evaluation project in which the state of some relevant throughput factor is not precisely known. In such a case, which is certainly not untypical of educational evaluation, the evaluator may have to base his or her judgments on a best guess as to the state of this part of the system. An important criterion for the quality of his or her judgments, then, is *robustness*. Generally, a judgment is robust if minor changes in the description of the state of the system do not lead to changes in the qualifications used in it. The idea underlying this criterion is obvious: Uncertainty about some part of the state of the system is less critical, the less dependent the qualifications are on the precise state the part of the system is in. The robustness of qualifications is usually assessed through a series of analyses in which changes in the values of some of the variables are made to simulate uncertainty about the state of the system, whereafter it is determined to what extent the qualifications would have to change. Obviously, robustness analyses are only possible if both the qualifications and the procedures leading to them are defined explicitly.

In the remainder of this paper, the results from a robustness study in a recent evaluation project in The Netherlands are reported to illustrate the possible contribution of robustness analysis to educational evaluation. The project was run by the Committee for the Evaluation of Elementary Education (CEB). In the next section, the problem addressed in the study is described. Subsequently, the methods of analysis will be given and the results will be discussed. The paper concludes with a discussion of the practical implications of the study.

INTRODUCTION TO THE PROBLEM

The evaluation committee was appointed by the Dutch Secretary of Education in 1991. Its mission was to evaluate the state of elementary education in the Netherlands from 1988-1992. In particular, the interest was in an evaluation of four different aspects of elementary education in this period, its level of achievements being one of them. The results of the evaluation were published recently (Commissie Evaluatie Basisonderwijs, 1994a, 1994b, 1994c, 1994d, 1994e). A fuller description of the assignment to the committee is given in Janssens (1995).

The committee had to report its findings at a level of aggregation that would suit a possible reorientation of the current policy of the Ministry of Education with respect to elementary education. Another constraint was that resources for data gathering were limited, and that the committee had to use existing sources of empirical data to perform its evaluation.

To present its evaluation of the achievements, the committee used the item material and scales from PPON. In this large-scale program for the assessment of educational progress in The

Table 1. Aggregation of PPON scales in evaluation project.

Subject	# Original Scales	# Aggregates
Dutch Language	13	7
Arithmetic	27	5
World Orientation	30	8
English	5	5
Traffic	2	1

Note. World Orientation is a combination of subjects. See Table 3.

Netherlands, which is run by the National Institute for Educational Measurement (Cito), the level of achievement in elementary education is periodically fathomed. The basic methodology used in PPON to scale the item pools and score the achievements is item response theory (IRT). The use of this methodology restricts the scaling of the items to the level of homogeneous subsets of the pool each measuring the same ability. An overview of the number of scales that were necessary to scale the item pools for the various subjects is given in Table 1.

For a complete review of the methodology used in PPON as well as reports of its assessments, the reader should consult van der Schoot (1993), Sijtstra (1992), Vinjé (1993), van Weerden (1993), Wijnstra (1998, 1990), and Zwarts (1990)

Definition of Qualifications

The selection of the qualifications by which the committee evaluated the achievements was guided by various considerations, three of which need to be explained here to be able to define the research problems addressed in this paper:

1. As already mentioned, the evaluation had to be reported at a level of aggregation suitable for recommendations on policy decisions. Therefore, it was necessary to combine sets of separate PPON scales into higher-level measures of achievement. For example, six separate scales for reading (Reading Reports; Reading Reflective Texts; Reading Persuasive Texts; Reading Maps; Reading References; and Reading Tables and Graphs) were combined into a single measure for Reading Comprehension. As IRT scales were not possible at this level of aggregation, the simple number of items correct score was used as a measure of achievement. However, this measure can be estimated from the scores on the IRT scales underlying the aggregate (see below). The number of aggregates in the evaluation is given in the last column of Table 1.
2. A second form of data reduction was also necessary to report the evaluations. The achievements of the population of students were in the form of distributions of scores. A usual way of defining qualifications for distinguishing between "good distributions" and "bad distributions" is in terms of their moments. Based on displays of the estimated distributions of the observed scores, the committee opted for qualifications for the first moments or means of the distributions. The main purpose of inspecting the displays was to get familiar with the relation between the location of the means and the shape of the left tails of the distribution. The qualifications were knowingly selected to be conservative; that is, relatively large proportions of students in the population had to be at the lower ends of the achievement scales before an unfavorable qualification applied.
3. Instead of qualifications in the form of a simple good/bad dichotomy, the committee chose three different qualifications for which the terms "Satisfactory" (Dutch: voldoende), "Moderate" (Dutch: matig) and "Unsatisfactory" (Dutch: onvoldoende) were used. As a compromise between the fact that evaluations in terms of observed scores are dependent on item pool content and the fact that a single set of qualifications is easier to communicate, the committee opted for a common definition of qualifications with adjustments for item pools that were deemed to be too difficult or too easy.

Table 2. Definition of qualifications used in evaluation.

Qualification	Mean of Score Distribution
Satisfactory	> 70%
Moderate	55% - 70%
Unsatisfactory	< 55%

Note. For item pools judged to be too difficult a downward adjustment of 10% and 5% was made for the lower bounds of Satisfactory and Moderate, respectively.

In fact, the definition of the qualifications was a long process in which such factors as familiarity with the curriculum, teaching practices, quality of the learning materials, previous evaluations, and extensive consulting of relevant parties played an important role. The results are given in Table 2.

Estimation of Mean Observed Scores

Two typical distributions of observed scores are given in Figure 1. Both distributions were estimated using the assumption of a correlation equal to .80 between the abilities on the underlying IRT scales.

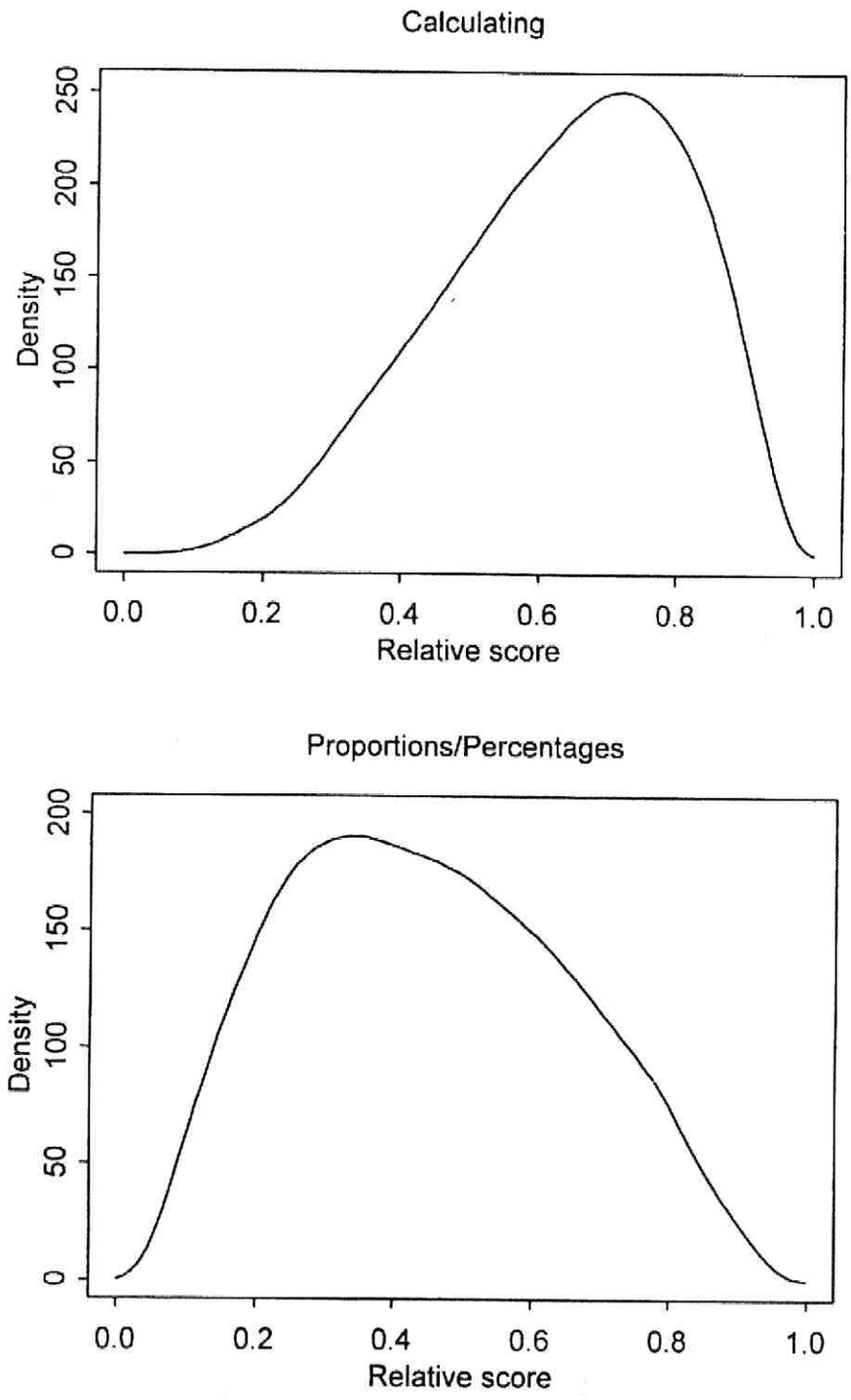


Fig. 1. Estimated observed-score distributions for: (a) Calculating; and (b) Proportions/Percentages.

The distribution for Calculating was evaluated as "Moderate". Its mean was just higher than the lower bound for this category but some 13% of the examinees solved less than one third of the items correctly. The distribution for Proportions/Percentages was estimated to have a mean in the category "Unsatisfactory". In this distribution, 36% of the examinees had less than one third of the items correct.

The means of the observed-score distributions were calculated from the item parameters estimated in the PPON projects. These estimates were obtained under the one-parameter logistic model with imputed values for the discrimination parameter (Verhelst, Glas & Verstralen, 1994). The ability distributions were scaled to be normal with mean 250 and standard deviation 50. Under the previous assumptions, the mean of an observed-score distribution can simply be calculated from the common marginal ability distribution and the sum of the response functions. This claim is proved in the Appendix.

Research Problems

The decision to use PPON item material and scales entailed two questions both related to the use of IRT in PPON.

First, though there is national agreement that the blueprints for the item pools had high content validity and that the sets of items in the pools covered the blueprints, some of the items were removed from the original pools in the scaling process. For example, for Arithmetic 4% of the items was removed from a pool of 491 items, whereas for Dutch 6% was removed from a pool of 498 items. These numbers were not large but important enough to pay attention to. As these items were removed on the basis of values of psychometric parameters and not of their content, it seems safe to conclude that:

1. The resulting pools still define the same ability variables, and that these variables have therefore not lost their validity; and
2. The removal of some of the items from the pools may nevertheless have had effects on the observed-score distributions, and hence on the judgments by the committee.

An important question is how serious these possible effects are.

Second, only the marginal ability distributions were available from PPON. As already explained, the choice for the mean as the critical moment of the distribution of observed-scores was based on plots of observed-score distributions. However, under the assumption of multivariate normality, to be able to plot observed-score distributions for aggregates of IRT scales, Pearson's correlation between the abilities must be known. (Remember that this requirement does not hold for the mean of the distributions.) As the abilities in each aggregate were "close", and numerous research projects have shown high correlations between subtests covering different aspects of, for example, language and arithmetic, the assumptions of correlations in the neighborhood of .80 seemed realistic. An important question is how serious the consequences of violation of this assumption are.

Both questions were addressed in a robustness study.

METHODS

Removal of Items

Four different procedures of item removal were simulated. In each procedure, after the removal of an item the mean of the observed-score distribution was calculated, and the correct qualification from Table 2 was selected.

The following procedures were studied:

1. Scaling

The pair of items with the smallest difference between their values for the difficulty parameter was selected, and one item of the pair was chosen at random and removed from the pool. The mean of the observed-score distribution was calculated, and the appropriate qualification was

identified The steps were repeated until the pool was empty. This procedure simulates item analysis in which the range of the scale values of the items has to remain maximal but redundancies are removed by eliminating items from subsets that cluster too strongly. The procedure applies when the ideal is a pool of items with uniformly distributed scale values.

2. Easy items
The item with the smallest value for the difficulty parameter was removed from the pool, the mean of the observed-score distribution was calculated, and the appropriate qualification was identified. The steps were repeated until the pool was empty. This procedure simulates the case where the item pool is considered too easy.

3. Difficult items
The previous procedure was repeated, but now at each step the most difficult item was removed.

4. Extreme items
This procedure is a combination of the previous two procedures. Alternately, the easiest and the most difficult item were removed. This procedure simulates the case where the item pool is considered to be on target but, for example, the distribution of abilities of the examinees is expected to have less spread than the item pool.

Correlation between Abilities
To assess the robustness of the observed-score distributions with respect to the correlation between the abilities, a Monte Carlo method was used to generate observed-score distributions on the sets of items in the aggregates for various values of the correlation coefficient. As a correlation between the abilities lower than .60 was most unlikely, the following values for the correlation coefficient were used: .60, .70, .80, and .90.

In the description of the Monte Carlo procedure below, the notation of the variables is the same as in the Appendix but the indices j = 1,...,J and i = 1,...,I are now used to denote the abilities and the items in a subset for the same ability, respectively:

1. For each simulated examinee, the values of the vector of abilities $(\theta_1,..,\theta_J)$ were drawn from a multivariate normal distribution with the assumed (common) value of the correlation coefficient.
2. The true scores $(t_1,...,t_J)$ were calculated as

$$t_j = \sum_{j=1}^{I} P_i(\theta_j), \quad j = 1, \ldots, J, \tag{1}$$

   and normed on [0,1].
3. The conditional distributions of X_j given $T_j=t_j$ are generalized binomial. Their probability functions, $Prob(X_j)$, were calculated using the first term in the expansion of the generalized binomial probability function given in Lord and Novick (1968, sect. 23.10).
4. The probabilities of the number-correct scores, $\sum_{j=1}^{J} X_j$, were calculated as

$$Prob\,(T=t) = \sum_{\Sigma X_j = t} \prod_j Prob\,(X_j). \tag{2}$$

The last step in the procedure made use of the fact that for a fixed examinee the observed scores X_j, j=1,...,J, were independent.

The accuracy of the approximation in Step 3 was checked against an algorithm suggested by Lord and Wingersky (1984) which produces the full generalized binomial distribution (see below).

The procedure was repeated for N=10,000 examinees. It should be noted, however, that for

each examinee not one realization of X_j given $T_j=t_j$ but its full conditional distribution was generated. The number is thus large enough to guarantee a smooth and stable result.

RESULTS

Graphs are used to present the results for the scaling procedure. In Figure 2, the mean observed

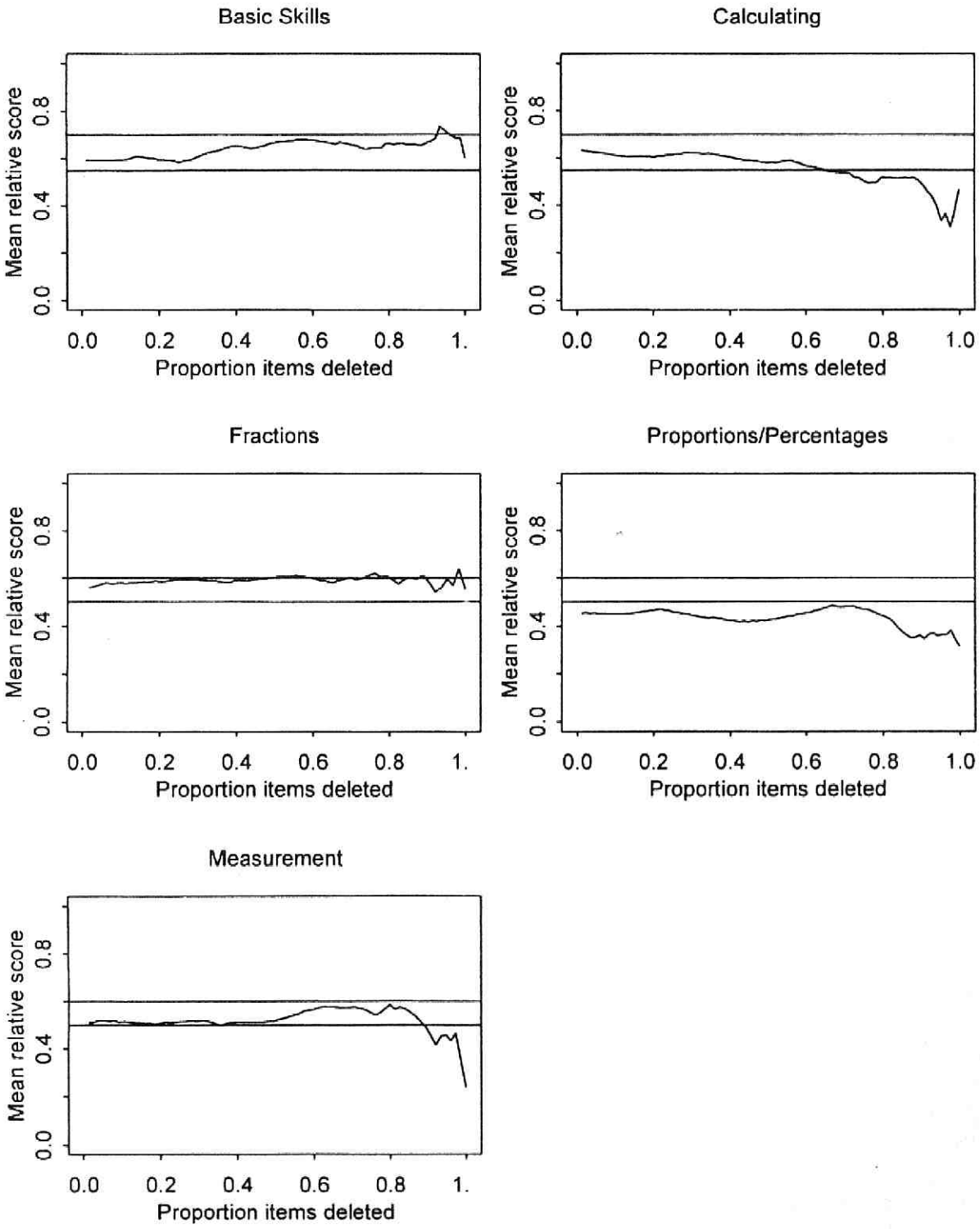


Fig. 2. Mean observed score as a function of the proportions of items removed due to scaling for Arithmetic.

relative scores for the five aggregates in Arithmetic are displayed as a function of the proportion of items removed due to scaling. The horizontal lines in the graphs denote the cut-off scores between the qualifications defined in Table 2. Generally, the curves follow a flat course, indicating extreme robustness of the mean with respect to the removal of items due to scaling. To cross one of the lines, the removal of 91% of the items for Basic Skills and 100% of the items for Proportions/Percentages was needed. For Calculating, the percentage was equal to 62%. The percentages for Fractions and Measurement are lower but still equal an impressive 45% and 33%, respectively. After these values, the two last curves started moving back and forth between the two sides of the upper (Fractions) and lower lines (Measurement). This behavior is typical of mean scores that were close to the borderline between two qualifications, remained there after removal of the items, but showed small fluctuations.

The results for the aggregates in the other subjects are given in Figures 3 through 6. The results are generally the same as for Arithmetic. All curves had a flat course, and, except for Reading English, at least 30-40% of the items had to be removed before the qualifications change. The case of Reading English is an interesting one. The curve was flattest of all curves in Figures 2-6, but the curve coincided with the upper line nearly perfectly. The same phenomenon

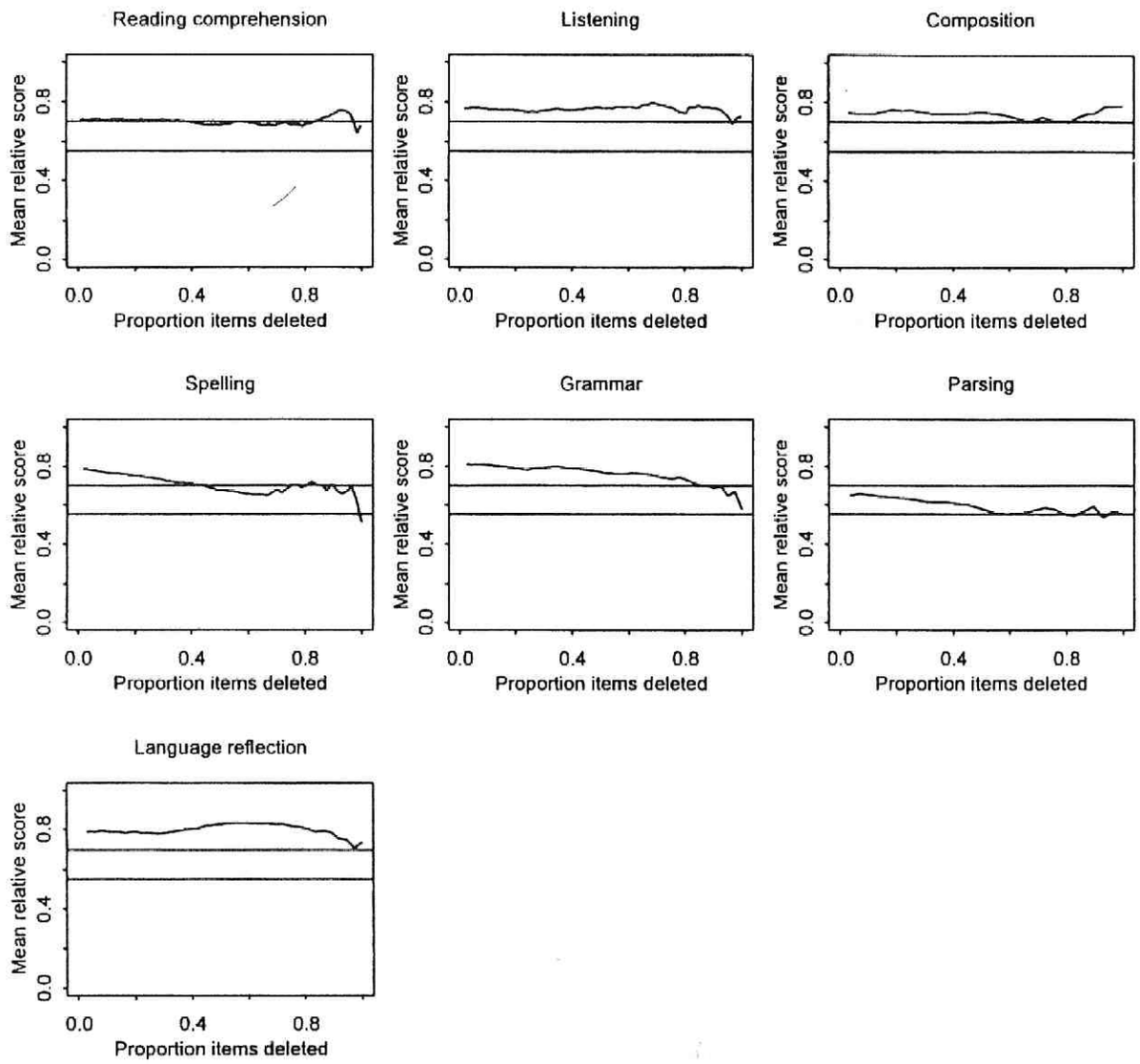


Fig. 3. Mean observed score as a function of the proportions of items removed due to scaling for Dutch.

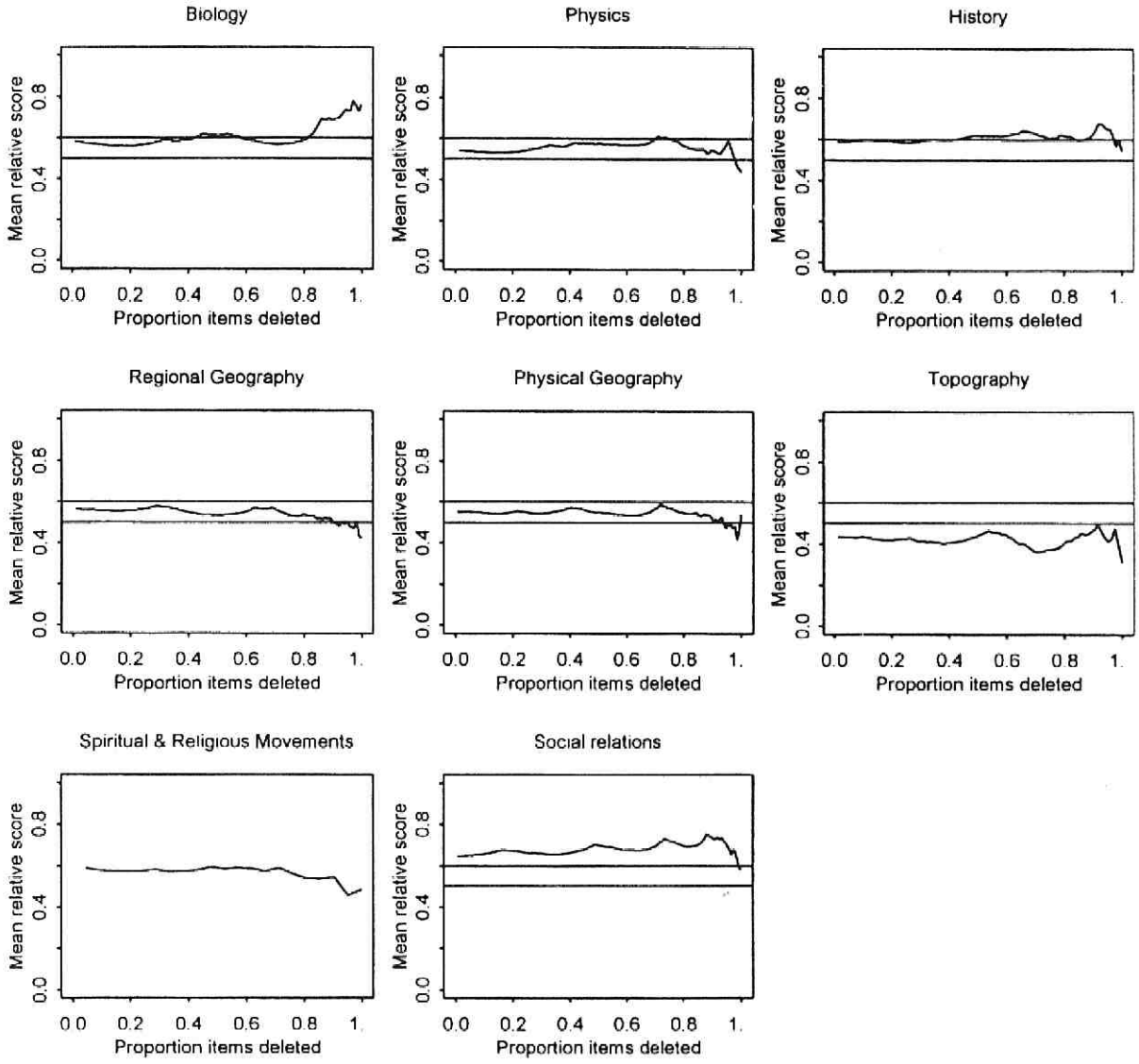


Fig. 4. Mean observed score as a function of the proportions of items removed due to scaling for World Orientation.

was observed for Reading Comprehension. Its curve was also flat and uniformly close to the line between "Satisfactory" and "Moderate". Nevertheless, 38% of the items had to be removed from the pool to change the qualification. At a later stage, the curve moved back to the original qualification. In its report, the committee made the provision that important parts of this aggregate were less favorable than the general impression suggested. Also, uncertainty was expressed due to the fact that data from an international comparison of achievements in Reading Comprehension had yielded conflicting information (Commissie Evaluatie Basisonderwijs, 1994a, sect. 5.1).

The results for all four principles of item removal are given in Table 3. The first column gives the percentages of items that had to be removed for the scaling procedure. The next three columns present the results for the other item removal procedures. Obviously, removal of the most difficult or easy items introduced a shift in the observed-score distributions, and generally the qualifications changed earlier than in the previous case. Nevertheless, with the exception of Measurement and Reading Comprehension for the removal of the easiest items and Biology for the most difficult items, the qualifications were remarkably robust for all aggregates. In these

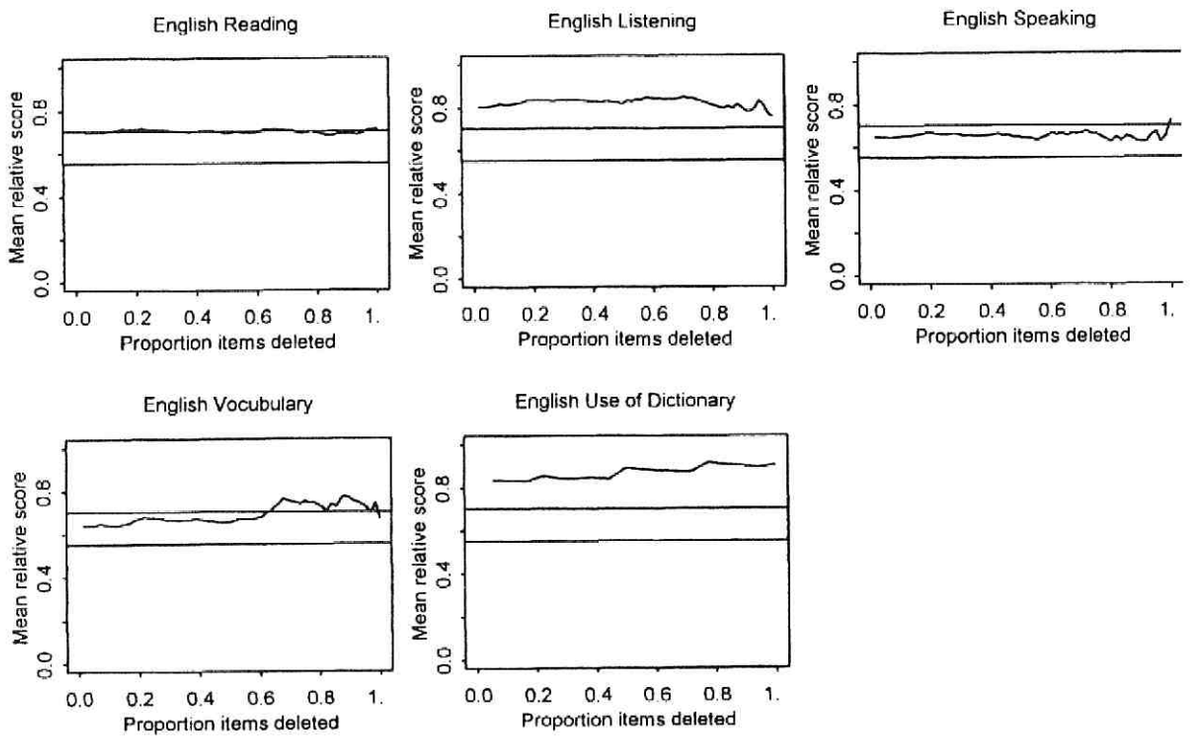


Fig. 5. Mean observed score as a function of the proportions of items removed due to scaling for English.

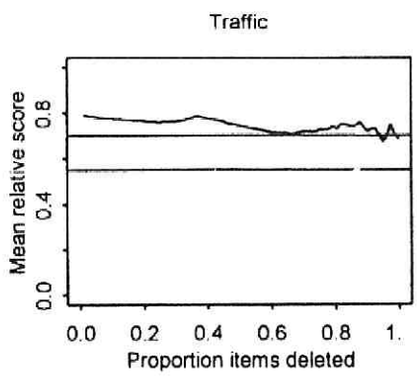


Fig. 6. Mean observed score as a function of the proportions of items removed due to scaling for Traffic.

exceptional cases of change, again the mean observed scores were already close to the borderline between two classifications for the intact item pool. For example, for Reading Comprehension the mean relative observed score for the intact item pool for the pool was .71, a result close to the cut-off score of .70 separating "Satisfactory" from "Moderate" (see Figure 2). The removal of the items with extreme difficulty values at both ends of the scale had, except for Reading of English, no noticeable effect on the qualifications. In the majority of the cases, nearly all items had to be removed before the qualification changed.

In Figure 7, two typical observed-score distributions for values of the correlation coefficient in the range from .60-.90 are shown. The effect of lowering the value of the correlation was a small shift of the mode of the distribution to the center of the scale. (However, remember that this phenomenon does not hold for the mean of the distribution. This parameter is independent

Table 3. Percentages of items needed to change the qualifications for the four methods.

Subject	Method			
	Scaling	Easy	Difficult	Extreme
Arithmetic				
Basic Skills	91	14	27	100
Calculating	62	29	16	100
Fractions	45	17	9	100
Proportions/Percentages	100	100	14	100
Measurement	33	3	25	27
Dutch				
Reading Comprehension	37	3	100	100
Listening	94	32	100	91
Composition	71	21	100	100
Spelling	39	34	100	100
Grammar	79	54	100	100
Parsing	71	37	13	100
Language Reflection	100	32	100	100
World Orientation				
Biology	41	28	4	26
Physics	66	13	17	100
Regional Geography	88	26	12	100
Physical Geography	91	16	17	100
Topography	100	100	17	100
History	40	47	100	100
Spiritual & Religious Movements	78	30	100	39
Social Relations	97	20	100	100
English				
Reading	3	3	100	6
Listening	96	41	100	100
Speaking	97	27	14	100
Vocabulary	59	24	13	100
Use of Dictionary	100	75	100	55
Traffic				
Practical Skills	91	42	100	100

of the value of the correlation coefficient.) Consequently, the value of the correlation coefficient does have some effect on the left tail of the distribution, but the effect is not dramatic. It seems safe to conclude that the relation between the mean and the left tail of the distributions observed by the committee does not change much in the neighborhood of r=.80.

As already observed, in Step 3 of the procedure for generating the observed-score distributions, an approximation to the generalized binomial distribution of X given T=t was made. The quality of the approximation was checked by comparing its results against those obtained for the exact distributions using the computer program AAPMOMT which implements the algorithm by Lord and Wingersky (1984) referred to earlier. The results were always virtually identical. Figure 8 gives the distributions for the same two aggregates as in Figure 7. The approximation proved to be excellent; the difference between the results of the two methods is hardly discernible.

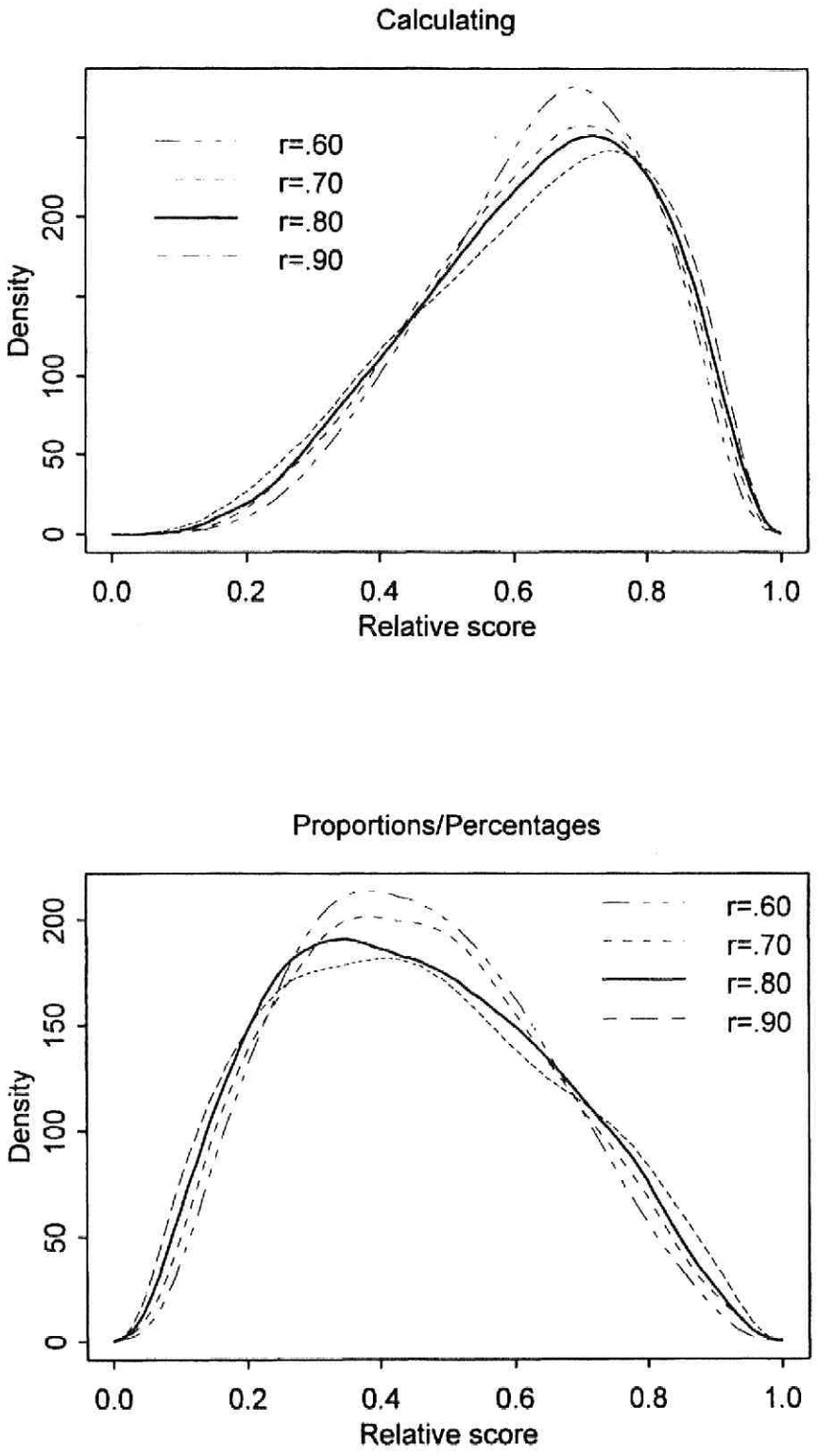
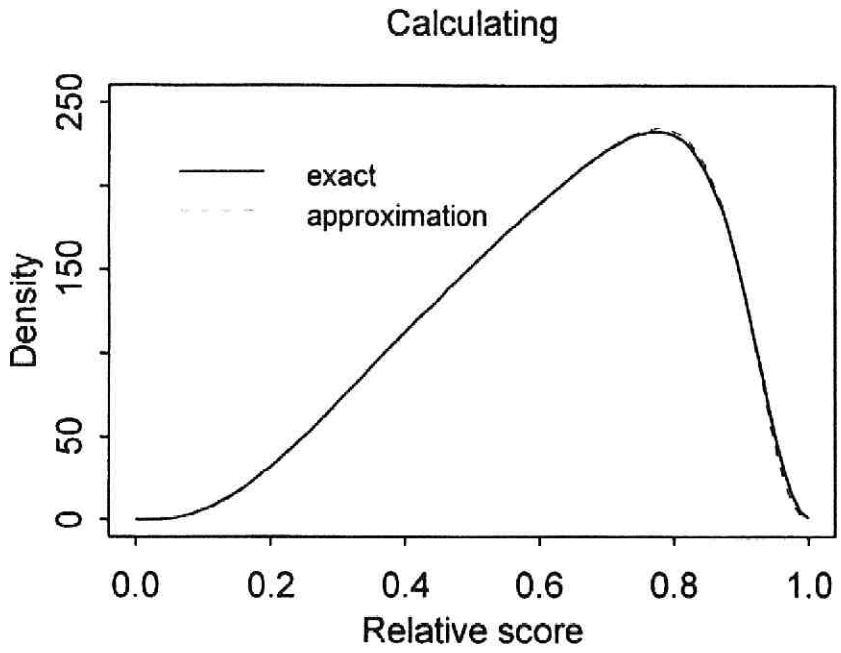


Fig. 7. Estimated observed-score distributions for: (a) Calculating; and (b) Proportions/Percentages (different correlation between abilities).



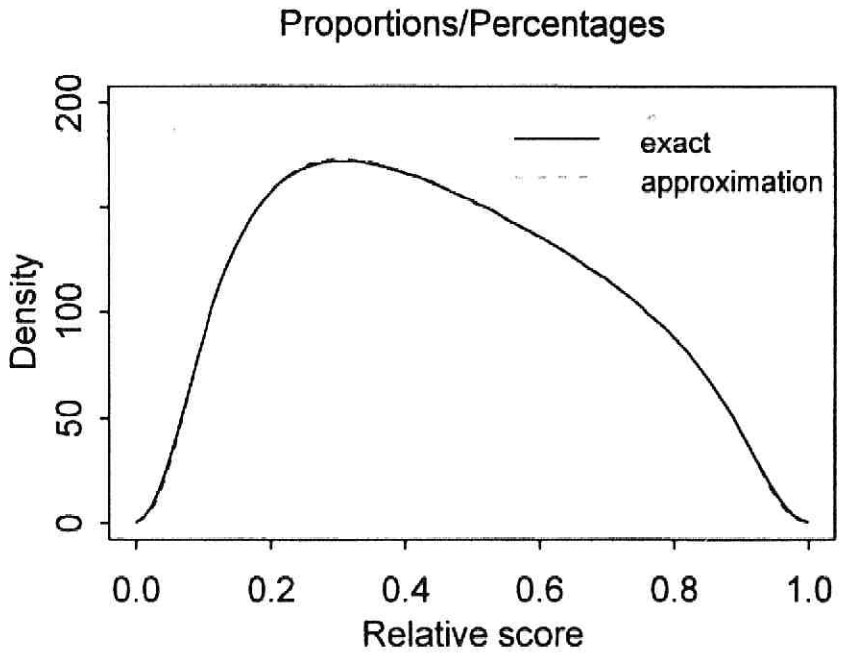


Fig. 8. Observed-score distributions estimated by: (a) Taylor approximation to generalized binomial; and (b) exact distribution function.

DISCUSSION

The main conclusion from the robustness study reported in this paper is that the qualifications used in the evaluation project are quite stable under the removal of items from the pool according to the four procedures defined above. Nearly all of the qualifications thus met a rigorous criterion of robustness.

In this study, the results for the scaling procedure are most important since this procedure comes closest to the procedure actually used in the PPON projects. However, it should be noted that the former is an idealized version of the latter, and that differences between the two may exist. Also, the procedure was applied to the item pools that were the results from PPON item analyses, and not to the original pools. Generalizing the findings to the original pools thus involves an element of extrapolation, albeit that the differences between the sizes of the two kinds of pools were generally small. Also, the fact that, with a few exceptions, remarkably robust results were obtained for procedures that deliberately made the item pools easier or more difficult does lend some support to the claim that this generalization is unlikely to involve serious bias.

It is emphasized that robustness of qualifications is only one necessary criterion which judgments in evaluation projects must meet, and that judgments are not automatically meaningful if they are robust. However, as illustrated in this paper, if uncertainty exists as to the knowledge base on which the judgments have to based, then robustness analysis is an excellent means to assess how serious the consequences of this uncertainty are.

APPENDIX

Independence of Mean Observed Score of Covariation between Abilities

For ease of exposition, the case of two distinct abilities is addressed. Let θ_1 and θ_2 be these two abilities. The bivariate distribution of the two abilities is represented by probability density function $f(\theta_1,\theta_2)$, whereas the marginal distributions of θ_1 and θ_2 are denoted as $f_1(\theta_1)$ and $f_2(\theta_2)$. Let X_1 and X_2 be the observed scores on the item sets measuring θ_1 and θ_2 and T_1 en T_2 the classical true scores for these observed scores.

In PPON, the marginal distributions of θ_1 and θ_2 are scaled to have common marginal densities:

$$f_1(\theta_1)=f_2(\theta_2)=f(\theta).$$

This feature is used in the proof below. The first step in the derivation follows from classical test theory, whereas the other steps are straightforward. Indices i and j denote items measuring the first and second ability, respectively. The proof runs as follows:

$$\begin{aligned}
E(X_1 + X_2) &= E(T_1 + T_2) \\
&= \int\int [\sum_i \boldsymbol{P}_i(\theta_i) + \sum_j \boldsymbol{P}_j(\theta_j)]\, f(\theta_1,\theta_2) d\theta_1 d\theta_2 \\
&= \int \sum_i \boldsymbol{P}_i(\theta_1) [\int f(\theta_1,\theta_2) d\theta_2] d\theta_1 + \int \sum_j \boldsymbol{P}_j(\theta_2) [\int f(\theta_1,\theta_2) d\theta_1] d\theta_2 \\
&= \int \sum_i \boldsymbol{P}_i(\theta_1)\, f_1(\theta_1) d\theta_1 + \int \sum_j \boldsymbol{P}_j(\theta_2)\, f_2(\theta_2) d\theta_2 \\
&= \int [\sum_i \boldsymbol{P}_i(\theta) + \sum_j \boldsymbol{P}_j(\theta)] f(\theta) d\theta.
\end{aligned}$$

Hence, when calculating the mean observed score, possible covariation between the underlying abilities can be ignored, and the item response function may be summed across abilities.

REFERENCES

American College Testing (1993). *AAPMOMT* (computer program). Iowa City, Iowa.

Commissie Evaluatie Basisonderwijs (1994a). *Inhoud en opbrengsten van het basisonderwijs* (Deelrapport 1). De Meern: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994b). *Onderwijs op maat* (Deelrapport 2). De Meern: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994c). *Onderwijs aan jonge kinderen* (Deelrapport 3). De Meern: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994d). *Onderwijs gericht op een multiculturele samenleving* (Deelrapport 4). De Meern: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994e). *Zicht op kwaliteit* (Eindrapport). De Meern: Inspectie van het Onderwijs.

Janssens, F.J.G. (1995). De evaluatie van het basisonderwijs door de CEB: aanpak en werkwijze. *Tijdschrift voor Onderwijsresearch, 20*, 3-12.

van der Linden (in press). Standards for standard setting in large-scale assessments. In *Proceedings of the Joint Conference on Large-Scale Assessments*. Washington, DC: National Assessment Governing Board & National Center for Education Statistics.

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

Lord, F.M. & Wingersky, M.S. (1984). Comparison of IRT true-score and equipercentile observed score "equatings". *Applied Psychological Measurement, 8*, 453-461.

van der Schoot, F. (1993). *Verantwoording van de peiling verkeerssituatie einde basisonderwijs* (PPON-report no.7). Arnhem: Instituut voor Toetsontwikkeling Cito.

Sijtstra, J. (1992). *Balans van het taalonderwijs halverwege de basisschool*. Arnhem: Instituut voor Toetsontwikkeling Cito.

Verhelst, N.D., Glas, C.A.W. & Verstralen H.H.F.M. (1994). *OPLM: The One-Parameter Logistic Model* (computerprogramma en manual). Arnhem: Instituut voor Toetsontwikkeling Cito

Vinjé, M. (1993). *Balans van het Engels aan het einde van de basisschool*. Arnhem: Instituut voor Toetsontwikkeling Cito.

Wijnstra (1988). *Balans van het rekenonderwijs in het basisonderwijs*. Arnhem: Instituut voor Toetsontwikkeling Cito.

Wijnstra, J. (1990). *Verantwoording van de rekenpeiling medio en einde basisonderwijs 1987*. Arnhem: Instituut voor Toetsontwikkeling Cito.

Zwarts, M. (1990). *Balans van het taalonderwijs aan het einde van het basisonderwijs*. Arnhem: Instituut voor Toetsontwikkeling Cito.

Manuscript ontvangen 9-8-1994
Definitieve versie ontvangen 7-11-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, pp. 28-33

De opbrengsten van het basisonderwijs volgens de CEB:
enkele kanttekeningen bij de gevolgde normeringsprocedures

Johan M. Wijnstra
Instituut voor Toetsontwikkeling (Cito), Arnhem

ABSTRACT

In 1993 the Committee for the Evaluation of Primary Education evaluated the outcomes of primary education in the Netherlands using data from the Dutch National Assessment Program in Education. In this article the committee's criteria are questioned from the principles of the assessment program and their practical value. Further, the implicit assumptions in aggregating subdomains into higher order domains are pointed out. It is concluded that in an evaluation of educational outcomes a sound, well-founded rationale for the criteria to be used is of the utmost importance.

INLEIDING

Het project Periodieke Peiling van het Onderwijsniveau (PPON) dat door het Instituut voor Toetsontwikkeling (Cito) wordt uitgevoerd in opdracht van de Minister van Onderwijs en Wetenschappen, heeft als opdracht beschrijvende rapporten over de stand van zaken in het basis- en speciaal onderwijs te produceren met betrekking tot het aanbod en de opbrengsten op basis van peilingen bij steekproeven van scholen en leerlingen. Het Cito is (tot nu toe) niet gerechtigd kwaliteitsuitspraken aan de gegevens te verbinden. De PPON-rapporten moeten aanknopingspunten bieden voor een (publieke) discussie over de inhoud en het niveau van het onderwijs.

De Commissie Evaluatie Basisonderwijs (CEB) heeft bij de beoordeling van de opbrengsten van het basisonderwijs uitvoerig gebruik gemaakt van PPON-gegevens (Commissie Evaluatie Basisonderwijs, 1994a; 1994b, verder aan te duiden als respectievelijk eindrapport en deelrapport 1). De CEB heeft twee normeringen op de gegevens toegepast, één gebaseerd op panelbeoordelingen in het licht van de kerndoelen die in 1993 in het Staatsblad zijn verschenen (Besluit kerndoelen basisonderwijs, 1993) en één op eigen gezag.

De panelbeoordeling, waarop de eerste normering is gebaseerd, bestond uit twee fasen. In eerste instantie is vastgesteld of de domeinbeschrijvingen waarop de PPON-toetsen zijn gebaseerd, voldoende overeenkomen met de kerndoelen. Vervolgens is de panels – bestaande uit inspecteurs, vakdeskundigen en ervaren leraren basisonderwijs – bij de relevante domeinen een oordeel gevraagd over de gewenste beheersing van de opgaven. De vraagstelling voor de panels bij de naar moeilijkheidsgraad geordende opgaven per domein luidde: *Welke opgaven moet een zwakke, respectievelijk goede leerling aankunnen, er van uitgaand dat de leerlingen onderwijs hebben gekregen dat door de kerndoelen is geïmpliceerd?* (Zwarts & Janssens, 1994). Deze procedure komt in essentie overeen met de procedures die eerder in het project zijn toegepast om bij verschillende groepen relevante respondenten (leraren basisonderwijs, ouderleden van medezeggenschapsraden van basisscholen en leraren uit verschillende typen voortgezet onderwijs) oordelen te verzamelen over het verwachte en gewenste vaardigheidsniveau van de gemiddelde leerling, met dien verstande dat in dit geval oordelen over de gewenste inhoudelijke beheersing op percentiel 25 (zwakke leerling) en 75 (goede leerling) werden gevraagd en dat de opgaven in het licht van de kerndoelen werden geplaatst. Net als in de PPON-procedures werd de definitie van een opgave aankunnen, afhankelijk van de soort vaardigheid, gedefinieerd als een succeskans van minimaal 80%, respectievelijk 50% volgens het toegepaste itemresponsmo-

Adres: Cito, Postbus 1034, 6801 MG Arnhem.

del. Uitgaande van de gegevens over het gewenste niveau van de zwakke leerling heeft de CEB vervolgens een norm toegepast op de discrepantie tussen het percentage leerlingen dat dit niveau haalt en het acceptabele percentage (25%). Deze norm leidt tot een oordeel met drie categorieën: onvoldoende, matig, voldoende.

Hoewel bij deze procedure ook wel kanttekeningen zijn te maken, bijvoorbeeld ten aanzien van de omvang van de panels, de procedures bij de vergelijking van de PPON-domeinbeschrijvingen met de kerndoelen en het gebruik van alleen de gegevens over het gewenste niveau van de zwakke leerling in de normering, kenmerkt deze procedure zich door een behoorlijke inhoudelijke verankering. Dat geldt voor de tweede procedure minder, waar ook verder de nodige kritiek op mogelijk is. In het vervolg van deze notitie worden bij deze procedure enkele kanttekeningen gemaakt, vanuit het perspectief of deze tot voldoende gefundeerde oordelen kan leiden.

DE NORMERING VAN DE COMMISSIE

De commissie heeft de opbrengsten van het vigerende onderwijs beoordeeld door te kijken naar de schattingen van de gemiddelde relatieve vaardigheidsscore per (sub)domein uit de verschillende peilingen. Vervolgens is daaraan een oordeel verbonden over de beheersing:

*Van een **voldoende** beheersing is sprake indien de populatie leerlingen gemiddeld genomen minimaal 70% van de opgaven goed maakt. Van een **matige** beheersing is sprake als gemiddeld minder dan 70%, maar meer dan 55% van de opgaven goed wordt gemaakt. Bij gemiddeld minder dan 55% is sprake van **onvoldoende** beheersing.* (Deelrapport 1, p. 55).

Op deze normeringsprocedure, die in een aantal gevallen naar beneden is bijgesteld, richt zich onze kritiek. De commissie geeft als rationale voor deze gekozen normen slechts de volgende formulering: *De Commissie acht deze normering redelijk en dus verantwoord.* (Deelrapport 1, p. 55).

Een dergelijke normering veronderstelt dat de gebruikte itemverzamelingen op de een of andere manier als representatief kunnen worden beschouwd voor de inhoud van het vigerende onderwijs en dat de normen passen bij de nagestreefde beheersing van deze stof. Deze veronderstelling is in elk geval niet te ontlenen aan PPON, hoewel de commissie schrijft: *De PPON-toetsen zijn zo geconstrueerd, dat leerlingen in principe de opgaven goed moeten kunnen maken.* (Deelrapport 1, p. 55). Deze zin is eenvoudig onjuist en in tegenspraak met de uitgangspunten bij de toetsconstructie, waarover we hier voor een goed begrip in het kort iets weergeven.

De gebruikte opgavenverzamelingen in de peilingen passen bij de daarvoor opgestelde domeinbeschrijvingen. De domeinbeschrijvingen bevatten een gestructureerde beschrijving van het mogelijke leer- en vormingsaanbod op een bepaald gebied en zijn niet beperkt tot het communale aanbod, enerzijds om te voorkomen dat de peilingen een conserverend effect hebben en anderzijds om ook op basis van peilingsresultaten nieuwe ontwikkelingen, die nog geen gemeengoed zijn, in de discussie te kunnen betrekken. Uit de peilingen van het aanbod blijkt ook dat de domeinbeschrijvingen in het algemeen meer omvatten dan enige school in het aanbod realiseert (het meest gedetailleerd is dit te vinden in de peiling wereldoriëntatie; vgl. Van Weerden, 1993). Bij de constructie en de selectie van de opgaven voor de peilingen gelden twee hoofdoverwegingen:

- de domeinbeschrijving moet zo volledig mogelijk gedekt worden;
- de toetsen en taken moeten een breed vaardigheidsbereik bestrijken.

Er is echter geen noodzaak de opgavenverzamelingen te beschouwen als steekproeven uit (oneindig) grote opgavenbanken die alle doelstellingen voor een domein met de gewenste moeilijkheidsgraad dekken. Door de toepassing van itemresponsmodellen kunnen zowel de vaardigheidsverdeling als de opgaven op dezelfde schaal worden afgebeeld, waarmee kan worden getoond wat typische leerlingen inhoudelijk kunnen en kennen. Door deze schalingsprocedure is er geen pertinente noodzaak te streven naar een bepaalde gemiddelde moeilijkheidsgraad of een bepaalde verdeling van de p-waarden van de opgaven en dat is ook nooit gebeurd. Als we al

ergens naar gestreefd hebben, dan hebben we eerder gemikt op een gemiddelde p-waarde van .50 dan .30 of .70 (vgl. Wijnstra, 1988, p. 153). Omdat de toetsen niet primair op deze basis zijn samengesteld, kunnen verschillen in geobserveerde moeilijkheidsgraad tussen de gebruikte toetsen voor verschillende domeinen zowel worden toegeschreven aan verschillen in beheersing als aan de toegepaste constructie- en selectieprocedures bij de samenstelling van de toetsen. Anders gezegd, voor een oordeel over het bereikte niveau moet de inhoud van de toetsen worden geraadpleegd. Op welke wijze de commissie tot een keuze voor de toegepaste normering is gekomen, wordt in deelrapport 1 nauwelijks verantwoord en dit komt de doorzichtigheid niet ten goede.

In dit verband is het ook de vraag wat de realiteitswaarde is van de toegepaste normering, in die zin dat men zich moet afvragen in welke orde van grootte de inspanning ligt om de kwalificatie 'voldoende' te bereiken in de gevallen dat de oordelen matig of onvoldoende luiden. De opbrengst van het reken/wiskunde-onderwijs wordt bijvoorbeeld over de gehele linie als matig of onvoldoende gekwalificeerd. In één geval blijkt dat het algemeen gemiddelde meer dan een halve standaarddeviatie zou moeten opschuiven om de kwalificatie 'voldoende' te bereiken. Effecten in deze orde van grootte worden bij door het onderwijs manipuleerbare variabelen zelden gevonden. Bij de door de commissie toegepaste normen is het reken/wiskunde-onderwijs wel bijna gedoemd om matig te blijven. Daarnaast kan de vraag gesteld worden hoe gevoelig de kwalificaties zijn voor veranderingen in de samenstelling van de opgavenverzamelingen. Door de commissie zijn de schattingen van de gemiddelde relatieve vaardigheidsscores gebruikt, ofwel de gemiddelde p-waarden, die afhankelijk zijn van de gebruikte opgavenverzamelingen. Hoe dichter deze waarden bij de grenswaarden in de normering liggen, hoe eerder de kwalificaties veranderen door opname of verwijdering van wat gemakkelijker, respectievelijk moeilijker opgaven (vgl. Van der Linden & Zwarts, 1995). Het is een bekend gegeven dat het beoordelen van de moeilijkheidsgraad van opgaven een notoir moeilijke zaak is. De commissie heeft voor dit probleem gestaan bij de keuze van de grenswaarden voor de normen in het algemeen en bij de keuze van deze 'standaardnormen' of de lichtere variant daarvan bij de afzonderlijke domeinen per leer- en vormingsgebied. Het is de vraag of de commissie bij een iets andere samenstelling van de opgavenverzamelingen dergelijke verschuivingen onderkend zou hebben.

Tabel 1. Resultaten van de beoordeling van de wereldoriënterende vakken*.

domein	opbrengst[1]	realisatie kerndoelen[1]
Natuur, waaronder biologie:		
Biologie	o	+
Natuurkunde	o	+
Geschiedenis	o	.
Aardrijkskunde:	+	
Regionale aardrijkskunde	o	+
Topografie	o[2]	-
Natuurkundige aardrijkskunde	o	+
Geestelijke stromingen	o	.
Maatschappelijke verhoudingen	+	.

* ontleend aan Deelrapport 1, p. 66;

[1] + = voldoende; o = matig; – = onvoldoende; . = onvoldoende aansluiting bij de kerndoelen volgens de panels

[2] norm waarschijnlijk op de verkeerde gegevens toegepast

Ter illustratie maken we enkele kanttekeningen bij de beoordeling van de opbrengst van het basisonderwijs op het gebied van wereldoriëntatie. De commissie-oordelen en de normering op de panelbeoordeling worden gepresenteerd in Tabel 1.

De commissie heeft bij de beoordeling de lage set normen gehanteerd (grenswaarden .50 en .60) met als magere toelichting: *De correctie wordt gemotiveerd door het feit dat de PPON-toetsen soms vrij moeilijk waren omdat ze vroegen naar stof die in eerdere leerjaren aan de orde was geweest en derhalve niet altijd verwacht mocht worden dat de leerlingen hun kennis nog volledig paraat konden hebben.* (Deelrapport 1, p. 66). Het resultaat daarvan is in de kolom 'opbrengst' vermeld. In de kolom 'realisatie kerndoelen' wordt het resultaat van de normering op de panelbeoordeling vermeld. Bij vergelijking van de twee normeringen volgt dan met een toon van verwondering: *Opvallend genoeg blijkt dat voor vrijwel alle leer- en vormingsgebieden het niveau dat de leerlingen nu behalen* ***niet hoger*** *(sic! bedoeld zal zijn: niet lager) is dan het niveau dat door de kerndoelen is geïmpliceerd. (...) Uitgaande van deze beoordeling is er geen noodzaak aanwezig om meer of anders aandacht te gaan besteden aan de wereldoriënterende vakken. Verondersteld moet dan ook worden dat de toetsen van PPON voor sommige onderdelen (te) moeilijk waren.*

Bij de laatste zin dringt zich natuurlijk ook de gedachte op dat de normering van de commissie nog aan de hoge kant was. Omgekeerd blijken de gemiddelde p-waarden van de opgavenverzamelingen voor de domeinen van wereldoriëntatie dicht bij de grenswaarden uit de normering te liggen (vgl. Zwarts & Janssens, 1994). Bij geschiedenis en biologie behoeven bijvoorbeeld slechts 4% van de moeilijkste opgaven uit de verzamelingen van ongeveer 100 opgaven te worden verwijderd om de kwalificatie 'voldoende' toe te kennen.

VERDERE AANNAMES BIJ DE NORMERING VAN DE COMMISSIE

In de rapportage van de peilingsuitkomsten worden over het algemeen rapportage-eenheden (schalen) van een beperkte omvang gehanteerd die didactisch betekenisvolle eenheden vormen, bijvoorbeeld bij rekenen/wiskunde basiskennis en begrip van procenten, handig procentrekenen en toepassingen met procenten in plaats van procenten. De commissie heeft in veel gevallen afzonderlijke schalen gegroepeerd tot bij de kerndoelen passende domeinen, bijvoorbeeld procenten en verhoudingen. Bij deze aggregatie kunnen verschillende kanttekeningen worden gemaakt die we toelichten aan de verdeling van de (genormaliseerde) moeilijkheidsparameters van de schalen uit het domein procenten en verhoudingen in Tabel 2 (in de PPON-rapporten worden de resultaten gerapporteerd op schalen met een gemiddelde van 250 en een standaarddeviatie van 50).

Om te beginnen zien we dat er aanzienlijke verschillen zijn in de verdelingen van de moeilijkheidsparameters tussen de schalen uit het kerndoelendomein. Bij de eerste rekenpeiling heeft geen proefafname plaatsgevonden (van deze open opgaven). De opgaven weerspiegelen de opvattingen van de toetsconstructeurs over de passendheid van de opgaven. Waarom er dan zulke grote verschillen tussen de schalen basiskennis en begrip, respectievelijk toepassingen optreden tussen procenten en verhoudingen is niet te zeggen. Het is alleen wel de vraag of de commissie hiermee bij de aggregatie niet op de een of andere manier rekening had moeten houden.

Per schaal is ongeveer een gelijk aantal opgaven gebruikt. Bij de aggregatie heeft de commissie feitelijk de afzonderlijke schalen gewogen met het aantal opgaven. De vraag is of dit terecht is. Hadden bijvoorbeeld de schalen basiskennis en begrip niet een veel hoger gewicht moeten hebben dan de toepassingenschalen en is het terecht dezelfde normen aan te leggen voor deze verschillende vaardigheden? Gelet op het verschil in de aard van de vaardigheid is bij de panelbeoordelingen – conform de suggestie van PPON – bij de toepassingenschalen uitgegaan van een succeskans van 50%, terwijl bij de andere schalen een succeskans van 80% is gehanteerd.

Tabel 2. Verdeling van de moeilijkheidsparameters in de schalen van het domein procenten en verhoudingen (aantallen opgaven).

	procenten			verhoudingen		
moeilijkheids-parameter	basisk./ begrip	handig %-rek.	toe-passingen	basisk./ begrip	toe-passingen	totaal
<= 100	-	-	-	1	-	1
101 - 125	-	-	-	2	-	2
126 - 150	-	-	-	1	-	1
151 - 175	1	-	-	1	-	2
176 - 200	1	2	-	-	-	3
201 - 225	2	2	-	9	3	16
226 - 250	5	4	2	3	2	16
251 - 275	5	3	2	-	4	14
276 - 300	1	7	4	1	4	17
301 - 325	1	1	5	-	-	7
326 - 350	1	-	1	-	3	5
351 - 375	-	-	-	-	-	-
376 - 400	-	-	2	1	-	3
> 400	1	-	1	-	-	2
totaal	18	19	17	19	16	89
gem. p	49	44	26	64	40	45

AFSLUITING

Dankzij de PPON-gegevens is er voor het eerst een brede waardering van de opbrengst van het basisonderwijs mogelijk. Zoals uit het voorgaande blijkt moet daarbij echter met de nodige voetangels en klemmen worden gerekend. Op zichzelf is het moedig van de CEB daartoe een poging te ondernemen, echter er zijn wel de nodige kanttekeningen bij te plaatsen die de oordelen van de commissie relativeren. Daarbij is met name het gebrek aan doorzichtigheid en de realiteitswaarde in het geding.

De CEB beveelt de minister aan te bevorderen *dat PPON in de toekomst mede uitspraken doet over de kwaliteit van de leerresultaten in het licht van de kerndoelen.* (Eindrapport p. 47). Anticiperend op de invoering van de kerndoelen is het PPON-projectteam al geruime tijd bezig alternatieve procedures te ontwerpen voor de verzameling van gegevens over het verwachte en gewenste niveau die een dergelijke beoordeling van de peilingsuitkomsten mogelijk moeten maken. Onze conclusie uit de verkenningen tot nu toe is dat het van het grootste belang is te zorgen voor een goed te verantwoorden inhoudelijke verankering van de normering, dat wil zeggen de opgaven op de verschillende schalen moeten gerelateerd worden aan (een vertaling van) de kerndoelen. Daarbij is het verder van belang dat de realiteitswaarde van de oordelen, respectievelijk de kerndoelen in het oog wordt gehouden. Dit is onder andere mogelijk door in de beoordelingsprocedure een of meer feedbackrondes in te bouwen.

LITERATUUR

Besluit kerndoelen basisonderwijs (1993). Staatsblad van het Koninkrijk der Nederlanden, nr. 264.

Commissie Evaluatie Basisonderwijs (1994a). *Zicht op kwaliteit.* Eindrapport. Z.p.: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994b). *Inhoud en opbrengsten van het basisonderwijs.* Deelrapport 1. Z.p.: Inspectie van het Onderwijs.

Linden, W.J. van der & Zwarts, M.A. (1995). Robustness of judgments in evaluation research. *Tijdschrift voor Onderwijsresearch, 20, dit nummer.*

Weerden, J. van (red.) (1993). *Balans van het wereldoriëntatie-onderwijs aan het einde van de basisschool.* Arnhem: Instituut voor Toetsontwikkeling.

Wijnstra, J.M. (red.) (1988). *Balans van het rekenonderwijs in de basisschool.* Arnhem: Instituut voor Toetsontwikkeling.

Zwarts, M.A. & Janssens, F.J.G. (1994). *Meten van kerndoelen met PPON-toetsen.* De Meern: Inspectie van het Onderwijs.

Manuscript ontvangen 17-6-1994
Definitieve versie ontvangen 24-10-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, pp. 34-41

De opbrengsten van het basisonderwijs: een repliek

W.J. van der Linden[1]
*Universiteit Twente**

M.A. Zwarts[1]
Inspectie van het Onderwijs

ABSTRACT

In this rejoinder to Wijnstra (1995), it is argued that, unlike the author seems to suggest, the definitions of the qualifications in the CEB evaluation project were not based on any assumptions regarding the distribution of difficulties in the item domains. Also, it is pointed out that statistical descriptions of the system of elementary education are contingent on the conditions under which the system exists, and that such descriptions are not sufficient to criticize the effects of the recommendations for change made by CEB. Finally, it is shown that the author's claim of aggregation of achievement scales being appropriate only for certain (unspecified) distributions of item difficulties ignores the unidentified character of the IRT difficulty scale, and hence does not apply.

Op 12 januari 1994 heeft de Commissie Evaluatie Basisonderwijs (CEB) zijn werkzaamheden voltooid met de aanbieding van een eindrapport en een viertal deelrapporten (Commissie Evaluatie Basisonderwijs, 1994a, 1994b, 1994c, 1994d, 1994e) aan de Minister van Onderwijs. Daarmee kwam een einde aan een intensieve werkperiode, die ruim twee jaar heeft geduurd. Het behoorde tot de opdracht aan de CEB om een zo breed mogelijke, evaluerende beschrijving van de toestand van het basisonderwijs te geven. Vervolgens moest deze evaluatie toegespitst worden aan de hand van een viertal specifieke vragen. Een volledig overzicht van de opdracht en de werkzaamheden van de CEB wordt gegeven in Janssens (1995).

Wijnstra (1995) heeft gelijk als hij in zijn notitie stelt dat de CEB bij zijn evaluatie van de opbrengsten van het onderwijs uitvoerig gebruik heeft gemaakt van de peilingen van de opbrengsten van het onderwijs door PPON. Op momenten als deze blijkt het belang van de beschikbaarheid van systematische empirische gegevens over de opbrengsten van het Nederlandse onderwijs. Gelukkig is het land dat goed peilingsonderzoek heeft.

Voor het betoog hieronder is het van belang om de precieze rol die de gegevens uit de PPON-projecten in de evaluatie van de CEB gespeeld hebben, scherp in het oog te houden. De CEB heeft de PPON-gegevens alleen gebruikt voor een *beschrijving* van de opbrengsten van het basisonderwijs. De *beoordeling* van deze opbrengsten is daarentegen volledig voor de rekening van de CEB. Bij de totstandkoming van deze beoordeling hebben ook deelbeschrijvingen uit andere bronnen voorgelegen (OVB-evaluatie; diverse SVO-projecten; dissertatie-onderzoek). Verder werd er door de CEB beoordeeld in het licht van evaluatieve uitspraken die eerder door anderen geuit waren. Sommige uitspraken zijn te vinden in de verslagen van de voornoemde projecten, terwijl andere werden geuit tijdens de raadpleging van zo'n 260 vertegenwoordigers van relevante instanties in het onderwijs en overige deskundigen. Ook werden in de PPON-projecten voor diverse leergebieden beoordelingsexperimenten uitgevoerd, waarin ouders, leraren BO en docenten VO hun verwachtingen en wensen ten aanzien van de opbrengsten kenbaar maakten. Naar het oordeel van de CEB was het algemene niveau van de verwachtingen en

[1] Beide auteurs waren respectievelijk als lid en stafmedewerker verbonden aan de Commissie Evaluatie Basisonderwijs.

* Adres: Universiteit Twente, Faculteit Toegepaste Onderwijskunde,. Postbus 217, 7500 AE Enschede.

Dit is het laatste artikel behorende bij het themanummer.

wensen onrealistisch hoog en waren ze in dit opzicht derhalve niet bruikbaar. Wel bruikbaar was het patroon in deze beoordelingen over de diverse leergebieden.

Ten slotte moet vermeld worden dat de CEB zijn beoordelingen heeft opgesteld op basis van een totaalbeeld van de toestand van het basisonderwijs dat onder andere berustte op een beschrijving van het aanbod, de leermaterialen, de pedagogisch-didactische aanpak, de leerlingen- en lerarenpopulatie alsmede van de fysieke, beleidsmatige en politieke omgeving van het onderwijs. De beschrijving van de opbrengsten vormde dus slechts een deel van dit totaalbeeld. Na de publikatie van de rapporten heeft zowel de pers als de politiek zich gestort op de constatering van de CEB dat het huidige aanbod overladen is. Voor de CEB waren andere constateringen, zoals die van het feit dat in het basisonderwijs, mede op grond van gebrek aan faciliteiten, de mogelijkheden van onderwijs op maat amper benut worden en dat er nog weinig systematisch aan kwaliteitszorg wordt gedaan, even belangrijk. Dergelijke constateringen vormen een onlosmakelijk geheel met de beoordelingen van de opbrengsten.

In zijn notitie maakt Wijnstra een aantal kanttekeningen bij de door de CEB gevolgde werkwijze, waar deze notitie een repliek op vormt. Zijn kanttekeningen laten zich groeperen rondom de volgende drie thema's:

1. de afhankelijkheid van de beoordelingen door de CEB van de toevallige psychometrische eigenschappen van de itemdomeinen;
2. de vraag of de normering door de CEB wel realistisch is geweest;
3. de problematiek van de aggregatie van de PPON-schalen.

ITEMDOMEINEN

In vrijwel alle peilingsonderzoek worden de opbrengsten van het onderwijs vastgesteld met behulp van itemdomeinen waaruit deelverzamelingen items aan verschillende steekproeven leerlingen worden voorgelegd. Het voordeel van deze werkwijze is dat de opbrengsten van het onderwijs met een maximale validiteit en nauwkeurigheid beschreven kunnen worden. Voorwaarde is wel dat de itemdomeinen omvangrijk en van een goede kwaliteit zijn. Om een hoge inhoudsvaliditeit te realiseren wordt een itemdomein in twee stappen geconstrueerd. In de eerste stap wordt het betreffende leer- of vormingsgebied volledig geclassificeerd. In de PPON-projecten staat deze stap bekend als de "domeinbeschrijving". In de tweede stap wordt een verzameling items geschreven die de gewenste verdeling over deze classificatie vertoont.

Het is de stellige indruk van de CEB dat zowel de domeinbeschrijvingen als de itemdomeinen uit de PPON-projecten een hoge inhoudsvaliditeit bezitten. Deze indruk is gebaseerd op kennisname van het materiaal en de lof die inhoudsdeskundigen aan PPON hebben toegezwaaid. Het resultaat is het gevolg van een werkwijze waarin bij de constructie van de opgaven de inhoudsvaliditeit voorop stond (Wijnstra, 1988, pag. 5). Uit commentaar van derden blijkt hoe de kwaliteit van de itemdomeinen voor bijvoorbeeld rekenen wordt gewaardeerd:

> "De PPON-rekentoetsen zijn een uitdrukking van een range van eindtermen: kennis en vaardigheden waarvan te verdedigen valt dat alle kinderen aan het einde van de basisschool ze zouden moeten beheersen' (Teunissen, 1988, pag. 169).
>
> "Nadere bestudering van de voorbeeldopgaven en van de totale collectie wijst uit dat de PPON-toetsen inderdaad het overgrote deel van de eindtermen bestrijken. ... Voorts kan worden vastgesteld dat de kwaliteit van de PPON-opgaven in het algemeen ruimschoots voldoende is om het onderwijsniveau passend te peilen" (Treffers, 1988, pag. 182).

Alleen Huitema (1988, pag. 164) maakt een opmerking waarin hij aangeeft dat hij meer opgaven had willen zien die minder van de Nederlandse taal afhankelijk zouden zijn en daardoor gemakkelijker waren. In het onderzoek van Zwarts en Janssens (1994), waaraan Wijnstra in de inleiding van zijn kanttekeningen refereert, gaf het panel van beoordelaars aan dat de toetsen voor rekenen goed samenvallen met de kerndoelen. Dezelfde conclusie werd getrokken voor Nederlandse taal, Engelse taal en verkeer. Alleen voor enkele domeinen van wereldoriëntatie moet er een voorbehoud worden gemaakt.

Naast de inhoudsvaliditeit wordt een tweede kenmerk van itemdomeinen gevormd door hun verdeling van de psychometrische eigenschappen van de items. Het is op dit punt dat de notitie van Wijnstra zich richt. Alvorens we de precieze inhoud van zijn kanttekening aangeven, is het van belang om op te merken dat in de PPON-projecten niet bewust naar een bepaalde verdeling van de psychometrische eigenschappen werd gestreefd. Dit gegeven wordt door Wijnstra verwoord wanneer hij stelt dat:

> "Door deze schalingsprocedure is er geen pertinente noodzaak te streven naar een bepaalde gemiddelde moeilijkheidsgraad of een bepaalde verdeling van de p-waarden van de opgaven en *dat is ook nooit gebeurd*" (pag. 4).

Zulk een streven zou in de praktijk ook nooit gerealiseerd hebben kunnen worden. Talloze empirische onderzoekingen laten namelijk zien dat, met uitzondering van enkele experimenten met matig succes voor onderdelen uit de wiskunde, het vrijwel onmogelijk is om verschillen tussen de eigenschappen van items die bij eenzelfde "cel" in een domeinbeschrijving behoren te voorspellen of te verklaren. Zulke items, die alle geschreven zijn bij eenzelfde onderwerp, op hetzelfde gedragsniveau en voor dezelfde leerlingen, blijken psychometrische verschillen op de leveren die in hoge mate afhangen van "toevallige", ogenschijnlijk irrelevante eigenschappen van de formulering van de items. De enige wijze waarop een gewenste verdeling van de psychometrische eigenschappen van de items kan worden verkregen is achteraf, door selectie aan de hand van empirische schattingen van deze eigenschappen. Wijnstra refereert aan deze bevindingen wanneer hij het volgende neerschrijft:

> "Het is een bekend gegeven dat het beoordelen van de moeilijkheidsgraad van opgaven een notoir moeilijke zaak is" (pag. 5).

We kunnen nu het door Wijnstra opgeworpen "meningsverschil" met de CEB scherp stellen. De CEB heeft, overeenkomstig de werkwijze van PPON, aangenomen dat de feitelijke verdeling van psychometrische eigenschappen in de itemdomeinen allereerst bepaald is door de opgestelde domeinbeschrijvingen en de gekozen verdeling van de items over de cellen van deze domeinbeschrijving. Gezien het toevalskarakter van de overblijvende verschillen tussen de items en de grootte van de geaggregeerde itemdomeinen waar de CEB mee heeft gewerkt, is vervolgens een beroep gedaan op de wet van de grote aantallen en werd aangenomen dat de feitelijk gecreëerde itemdomeinen psychometrisch een afspiegeling vormen van de domeinen aan mogelijk items die men zich bij de domeinbeschrijvingen in kan denken. Onder deze aanname is een benadering mogelijk die bekend staat als "domain-referenced measurement" en waarbij normen geformuleerd kunnen worden in termen van de aantal-goed scores van de populatie leerlingen voor het totale domein (zie bijvoorbeeld Hively, Maxwell, Rabehl, Sension & Lundin, 1973). De CEB heeft deze benadering gekozen; voor een definitie van de normen raadplege men Van der Linden en Zwarts (1995). Deze normering is uiteraard niet blind geweest en berust op een inhoudelijke beoordeling van alle beschikbare PPON publikaties en materialen. Voor een tweetal onderdelen van rekenen en taal alsmede voor het gehele gebied van wereldoriëntatie, bestond de overtuiging dat de hiervoor omschreven aanname niet volledig correct was en is door de CEB een correctie op de normen toegepast.

Het merkwaardige is nu dat Wijnstra de hiervoor beschreven werkwijze in de PPON-projecten nog eens herbevestigt, maar vervolgens suggereert dat de normering van de CEB alleen maar zin heeft als er binnen PPON *wel* naar een vooraf bepaalde verdeling van de moeilijkheden van de items in de itemdomeinen zou zijn gestreefd. Aan welke voorwaarden voldaan zou moeten worden en hoe deze "bepaalde" verdeling van itemmoeilijkheden er uit zou moeten zien, wordt evenwel niet meegedeeld. Op dit punt moet er echt sprake zijn van een misverstand. De werkwijze van de CEB berust niet op een aanname dat alle PPON-items een p-waarde in de buurt van .70 zouden bezitten of iets dergelijks. Zoals hierboven geschetst werd, is het enige dat gevraagd wordt dat de itemdomeinen representatief zijn voor de leerstof.

Het is wel zo dat men in tweede instantie, door itemanalyse op grond van empirische gegevens, de samenstelling van de oorspronkelijke domeinen kan wijzigen. Door PPON is langs deze weg een aantal items uit de oorspronkelijke domeinen verwijderd. Dit aantal was telkens klein. Het doel bij deze verwijdering was om schalen te creëren met een maximale breedte aan

waarden voor de moeilijkheidsparameter uit het gehanteerde IRT model. Zo werden bijvoorbeeld enkele items verwijderd die redundant waren omdat ze deel uitmaakten van een cluster van items met ongeveer dezelfde parameterwaarden. Ondanks de kleine aantallen waarom het gaat, voor rekenen bedraagt het aantal zo'n 4% van het oorspronkelijke domein, zou deze itemselectie de beoordelingen van de CEB hebben kunnen verstoren. Daarom is door de CEB voor alle zekerheid een robuustheidsanalyse uitgevoerd waarin deze itemselectie werd gesimuleerd en de gevolgen nagegaan werden. Voor een volledig overzicht van de resultaten van deze analyse wordt verwezen naar de figuren 3-6 in Van der Linden en Zwarts (1995). Het vlakke verloop van bijna al deze figuren over de gehele range van 0-100% wijst op een extreem hoge robuustheid van de beoordeling van de CEB ten opzichte van het gevolgde type itemanalyse voor de itemdomeinen in de PPON-projecten. Voor een groot aantal leergebieden moeten bijna alle items verwijderd worden alvorens het oordeel van de CEB zich dient te wijzigen. Alleen in een paar gevallen, waarin de gemiddelde geobserveerde scores voor de populatie leerlingen reeds dicht tegen de normen van de CEB aanzaten, kan er voor kleinere percentages verwijderde items een verandering van de beoordeling nodig zijn, die soms voor grotere percentages weer ongedaan gemaakt moet worden.

REALISTISCHE NORMERING?

Wijnstra vraagt zich hardop af wat de realiteitswaarde is van de door de CEB toegepaste normering. Terecht meet hij de betekenis van kwalificaties als "voldoende", "matig" en "onvoldoende" af aan de inspanning in het onderwijs die geleverd moet worden om de geobserveerde scoreverdeling naar een hogere kwalificatie op te laten schuiven. Deze operationele definitie was ook degene van de CEB en is er de reden van dat, zoals eerder reeds werd opgemerkt, de aanbevolen maatregelen en de beoordelingen niet los van elkaar gezien mogen worden. De kern van de opmerkingen van Wijnstra is dat de beoordelingen van de CEB effecten van maatregelen veronderstellen die zelden gevonden worden bij door het onderwijs manipuleerbare variabelen.

Met deze effecten en variabelen doelt Wijnstra ongetwijfeld op de door PPON gepubliceerde resultaten van analyses waarin de gemiddelde scores op de PPON-schalen gerelateerd worden aan achtergrondvariabelen als de gehanteerde leerboeken, geslacht, leeftijd, gewicht van de leerlingen in de formatiebepaling, e.d. De constatering dat in PPON effecten van de orde van grootte van 0,5 standaarddeviatie zelden geschat konden worden, is terecht. Voor bijvoorbeeld rekenen worden effecten van deze grootte vrijwel alleen aangetroffen, maar dan onmiddellijk voor een ongewenst groot aantal PPON-schalen, voor de gewichtscategorieën voor leerlingen zoals deze in de huidige financiering van het basisonderwijs gelden.

Door de in PPON gevonden effecten als norm voor onderwijsverandering te kiezen, maakt Wijnstra een klassieke fout in de interpretatie van de resultaten van statistische resultaten. Regressie van de resultaten van peilingsonderzoek op achtergrondvariabelen levert alleen een *beschrijving* op van de samenhang tussen variabelen onder de *huidige voorwaarden* in het onderwijs. De gedachte dat men daarmee causale verbanden vastgesteld heeft, die de effecten van onderwijsverandering voorspelbaar maken, is onterecht. Dergelijke verbanden kunnen nooit via peilingsonderzoek vastgesteld worden. Technisch gezien gaat het hier om het onderscheid tussen regressievergelijkingen die worden geschat in een beschrijvend onderzoek en differentiaalvergelijkingen die worden geverifieerd via experimenteel onderzoek. Tegen overspannen verwachtingen van het gebruik van regressie-analyse in peilingsonderzoek, werd reeds gewaarschuwd door Freudenthal (1975) in zijn bekende kritiek op de eerste internationale peilingen door de IEA, die in zijn rapporten dezelfde status aan regressie-analyse toeschreef.

In de tweede plaats is er genoeg empirisch onderzoek dat laat zien dat effecten van instructie- en schoolvariabelen op de leerprestaties groter dan 0,5 standaarddeviatie wel degelijk tot de mogelijkheden behoren. In een recent overzicht van Fraser (1989, Table 2.1) heeft maar liefst een derde van de bestudeerde variabelen een effectgrootte die boven deze norm ligt.

Er moet dus rekening mee gehouden worden dat de CEB zich in zijn aanbevelingen op een

grondige herziening van een aantal voorwaarden in het onderwijs richt. Zoals al gememoreerd werd, is een behoorlijke inperking van het verplichte aanbod één van de aanbevelingen, zodat er per saldo veel meer tijd voor minder onderwerpen zal ontstaan. Nu blijken te grote percentages scholen niet eens aan de behandeling van sommige onderwerpen toe te komen. Ook het vrijmaken van extra middelen om tijdens de eerste vier jaren van het basisonderwijs vroegtijdige diagnose te kunnen vervolgen met een aanbod aan onderwijs op maat, is een krachtige maatregel. Voeg daarbij nog eens het pleidooi voor een opbrengstgerichte cultuur, waarin kwaliteitszorg een vast gegeven is, dan is het niet overdreven om een pittige verhoging van de opbrengsten te verwachten. Om deze voorspelling te concretiseren het volgende. De curve voor Proportions/ Percentages in figuur 1 in Van der Linden en Zwarts (1995) heeft geleid tot het oordeel "onvoldoende". Het is helemaal niet onrealistisch om aan te nemen dat, waar nu zelfs gemiddeld zo'n 8-9% van de leerlingen überhaupt geen instructie in deze onderwerpen krijgt (Wijnstra, 1988, par. 5.7), de bovengenoemde maatregelen de curve een vorm aan kunnen doen nemen die dicht bij die van de eerste curve in deze figuur ligt, die de beoordeling "matig" vertegenwoordigt.

Overigens komen de beoordelingen van de CEB, op een enkele uitzondering na, vrijwel perfect overeen met die van het PPON-team in een recente aflevering van een themanummer van het blad *School* (1994). Dit themanummer verscheen gelijktijdig met de rapporten van de CEB en is onafhankelijk tot stand gekomen. Hieronder volgt een selectie van de evaluatieve uitspraken door het PPON-team, waarvoor we de lezer oproepen om deze met de beoordelingen in de CEB-rapporten te vergelijken:

"Als goed uit de bus komen de prestaties van leerlingen bij spellen en technische lezen. Als zwak: begrijpend lezen, schrijven en spreken." (pag. 24-25)

"Het is waar dat de kwaliteit van het schriftelijk werk in groep 8 soms erg tegenvalt. Het wonderlijke is dat we in groep 5 vergelijkbare opdrachten hebben gegeven, die goed werden uitgevoerd." (pag. 25)

"De resultaten op het onderdeel cijferen zijn goed te noemen: veel leerlingen beheersen de standaardalgoritmen voor de bewerkingen optellen, aftrekken, vermenigvuldigen en delen met gehele getallen." (pag. 28)

"De resultaten op het onderdeel procenten vallen tegen. Procentopgaven worden slechts door een kleine groep leerlingen goed beheerst. Een gemiddelde leerling heeft geen goed begrip van wat procenten zijn en hij is niet in staat eenvoudige procentberekeningen met een grote kans op succes uit te voeren." (pag. 29)

"Hoofdrekenen moet beter, schatten is onvoldoende, maar dat is ook nog vrij nieuw in het onderwijs. Inzichten in het fundament van breuken, verhoudingen en procenten moet veel beter. De basiskennis van procenten blijkt ook onvoldoende. In sommige methodes wordt er pas in groep 8 aandacht aan besteed. Er is ruimte om die lacunes te bestrijden, we cijferen immers te veel." (pag. 30)

"Bij taal vond ik het zeer opvallend dat de produktieve vaardigheden, stellen en spreken, zwaar beneden de verwachtingen bleven." (pag. 39)

"Spellen bleek boven verwachting goed." (pag. 39)

"Bij rekenen is het algemene beeld dat breuken, procenten en verhoudingen er mager uitkomen." (pag. 39)

AGGREGATIE VAN SCHALEN

In de PPON-projecten wordt gerapporteerd op een hoog niveau van detaillering. Bijvoorbeeld voor de opbrengsten van het rekenen aan het einde van het basisonderwijs werden de opbrengsten gepeild met behulp van 27 verschillende schalen. Het aantal schalen is mede het resultaat van het gebruik van itemresponstheorie (IRT), dat aan de schaling van de gebruikte opgaven de eis van eendimensionaliteit oplegt. Dit is een duidelijk winstpunt. We weten daardoor welke delen van leergebieden we als homogeen kunnen beschouwen en kunnen bijvoorbeeld bij de herziening van didactische methoden met dit gegeven rekening houden.

Wie naar het beleid rapporteert, is echter onderhevig aan de wetten die daarbij behoren. Je betreedt dan de wereld van de *executive summaries*, moet aansluiten bij de "lopende discussies" en vooral "heldere boodschappen" brengen. Voor de rapportage van de opbrengsten van het basisonderwijs ligt dat niet anders. Een beleidsmaker heeft er niet zoveel aan om te weten of het nu met "vermenigvuldigen" of met "delen" fout zit. Hij is meer geïnteresseerd in de opbrengsten voor de "basisbewerkingen" en wil antwoord op de vraag of daar in het onderwijs nu meer of minder aandacht aan geschonken moet worden dan aan bijvoorbeeld "toepassingen". Kortom, er is aggregatie van de gegevens nodig tot het niveau van de grotere onderdelen van de leergebieden–ongeveer het niveau dat gebruikt werd in de uitspraken van het PPON-team in de citaten hierboven. Dat is niet erg. Er gaat wel wat informatie verloren, maar er wordt aan effect gewonnen. En voor achtergrondinformatie of in discussies met vakgenoten kan altijd weer teruggegaan worden naar het oorspronkelijke niveau aan detaillering.

In de CEB-rapporten is welbewust gekozen voor een indelingsniveau dat relevant is voor de maatschappelijke discussie over de invoering van de kerndoelen in het basisonderwijs. Deze keuze brengt onherroepelijk met zich mee dat voor de scores een andere schaal gekozen moet worden dan de IRT-schalen die in de PPON-projecten werden gehanteerd. Door de CEB zijn daarom alle berekeningen gedaan op de schaal van de aantal-goed scores voor geaggregeerde itemdomeinen uit de PPON-projecten. Voor de berekening van de populatieverdelingen op deze schaal is dankbaar gebruik gemaakt van de schatting van de IRT-parameters van de items uit de PPON-projecten. Voorbeelden van deze berekeningen zijn te vinden in Van der Linden en Zwarts (1995).

Wijnstra maakt twee opmerkingen bij de gevolgde aggregatieprocedure. Allereerst stelt hij de vraag of een ongeveer gelijke weging van de afzonderlijke PPON-schalen wel terecht is en de opbrengsten op de schalen voor basiskennis en -begrip niet zwaarder gewogen zouden moeten worden dan de opbrengsten op de meer toepassingsgerichte schalen. Ten tweede presenteert hij in tabel 2 de verdelingen van de waarden voor de moeilijkheidsparameter uit de IRT voor de onderdelen procenten en verhoudingen en vraagt zich af "of de commissie hiermee bij de aggregatie niet op de een of andere manier rekening had moeten houden" (pag. 7).

Op de eerste vraag is onze reactie kort. De feitelijk gehanteerde weging verdient de voorkeur. De verdeling van de aantal-goed scores op geaggregeerde domeinen leidt tot totaaloordelen aan de hand van een omnibusmaat. Bij ongelijke weging zouden deze totaaloordelen zeer gevoelig worden voor de opbrengsten op slechts enkele van de oorspronkelijke schalen en zouden de opbrengsten op de andere schalen dus noodzakelijk delen in het oordeel dat de eerste afdwingen. Bij gelijke weging kan dezelfde gevoeligheidsproblematiek reeds spelen als één van de onderliggende vaardigheden of kennisgebieden in sterke mate van de andere afwijkt. Ook dit is ongewenst, maar onvermijdelijk. In zo'n geval heeft de CEB in zijn rapporten de toelichting op de tabellen gebruikt om deze afhankelijkheid zichtbaar te maken. Zie hiervoor bijvoorbeeld de toelichting op de rol die het lezen van rapporterende en argumentatieve teksten speelt in het totaaloordeel over de opbrengsten voor begrijpend lezen (Commissie Evaluatie Basisonderwijs, 1994a, pag. 58).

De reactie op de tweede vraag is methodologisch van aard. Vergelijking van de spreidingen van de twee verdelingen van de moeilijkheidsparameters voor procenten en verhoudingen in tabel 2, zoals Wijnstra doet, is zinloos. In feite worden hier appels met peren vergeleken. Onder het in PPON gebruikte IRT model hebben deze verdelingen namelijk ieder een willekeurige eenheid en nulpunt (zie bijvoorbeeld Hambleton & Swaminathan, 1985, par. 4.2). We kunnen de spreiding van de parameterwaarden op de individuele PPON-schalen zo breed of smal maken als we zelf willen, zonder hierdoor de passing van het model aan te tasten. Ieder van deze spreidingen zou ook *dezelfde* verdeling van de aantal-goed scores voor het betreffende itemdomein opleveren en derhalve tot *hetzelfde* oordeel van de CEB hebben geleid. Op de door Wijnstra opgeworpen vraag bestaat geen antwoord, omdat ze geen probleem aansnijdt.

DISCUSSIE

In deze reactie is betoogd dat er geen meningsverschil bestaat over de wijze waarop in de PPON-projecten de itemdomeinen zijn geconstrueerd. Volgens Wijnstra is alleen gestreefd naar een inhoudelijke dekking en niet naar een bepaalde verdeling van psychometrische eigenschappen van de items, en dat is precies de veronderstelling waaruit de CEB vertrokken is. Merkwaardig blijft het betoog van Wijnstra dat de werkwijze en normering van de CEB alleen zin hebben als de moeilijkheid van de items een bepaalde verdeling zou bezitten, waarvan hij de aard overigens niet omschrijft. De verwijzingen van Wijnstra naar de regressie van de PPON-resultaten op een aantal achtergrondvariabelen en de betekenis van de verdeling van de moeilijkheidsparameters in de itemdomeinen berusten beide op een methodologisch misverstand, waarvan de aard door ons is aangegeven. Blijft staan de vraag of de combinatie van beoordelingen en aanbevelingen door de CEB realistisch is geweest. Het deed ons deugd te constateren dat het PPON-team in een onafhankelijke publikatie deze beoordelingen vrijmoedig bijviel.

Het was niet de bedoeling van deze reactie om de betekenis van peilingsonderzoek voor het Nederlandse onderwijs ter discussie te stellen. Deze staat buiten kijf. In de aanhef is al gesteld dat op momenten als de evaluatie door de CEB het belang van systematisch empirisch onderzoek naar de opbrengsten van het onderwijs blijkt. In zijn nabeschouwingen heeft de commissie het belang van peilingsonderzoek onderstreept en ervoor gepleit om de systematiek van peilingsonderzoek uit te breiden naar een periodieke verzameling van empirische gegevens over aanbod, context en processen in het onderwijs. Ook is een systematische vergelijking van peilingsresultaten in internationaal verband gewenst (Commissie Evaluatie Basisonderwijs, 1994e)

LITERATUUR

Commissie Evaluatie Basisonderwijs (1994a). *Inhoud en opbrengsten van het basisonderwijs* (Deelrapport 1). De Meern: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994b). *Onderwijs op maat* (Deelrapport 2). De Meern: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994c). *Onderwijs aan jonge kinderen* (Deelrapport 3). De Meern: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994d). *Onderwijs gericht op een multiculturele samenleving* (Deelrapport 4). De Meern: Inspectie van het Onderwijs.

Commissie Evaluatie Basisonderwijs (1994e). *Zicht op kwaliteit* (Eindrapport). De Meern: Inspectie van het Onderwijs.

Fraser, B.J. Research syntheses on school and instructional effectiveness. *International Journal of Educational Research, 13*, 707-720.

Freudenthal, H. (1975). Pupils' achievements internationally compared: The IEA. *Educational Studies in Mathematics, 6*, 127-186.

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff.

Hively, W., Maxwell, G., Rabehl, G., Sension, D. & Lundin, S. (1973). *Domain-referenced curriculum evaluation: A technical handbook and a case study from the MINNEMAST project*. Los Angeles, CA: Center for the Study of Evaluation, University of California.

Huitema, S. (1988). We overvragen de basisschool. In J.M. Wijnstra (red.), *Balans van het rekenonderwijs is de basisschool* (PPON-reeks nr. 1). Arnhem: Instituut voor toetsontwikkeling Cito.

Janssens, F.J.G. (1995). De evaluatie van het basisonderwijs door de CEB: aanpak en werkwijze. *Tijdschrift voor Onderwijsresearch, 20*, 3-12.

Van der Linden, W.J. & Zwarts, M.A. (1995). *Tijdschrift voor Onderwijsresearch, 20*, 13-27.

School (1994). *Een beeld van een basischool (themakatern). 22*, 19-40.

Teunissen, F. (1988). Een hoge norm. In Wijnstra, J.M. (red.), *Balans van het rekenonderwijs is de basisschool* (PPON-reeks nr. 1). Arnhem: Instituut voor toetsontwikkeling Cito.

Treffers, A. (1988). Over de merkbare invloed van onderwijsmethoden op leerprestaties. In Wijnstra, J.M. (red.), *Balans van het rekenonderwijs is de basisschool* (PPON-reeks nr. 1). Arnhem: Instituut voor toetsontwikkeling Cito.

Wijnstra, J.M. (red.) (1998). *Balans van het rekenonderwijs is de basisschool* (PPON-reeks nr. 1). Arnhem: Instituut voor toetsontwikkeling Cito.

Wijnstra, J.M. (1995). De opbrengsten van het basisonderwijs volgens de CEB: enkele kanttekeningen bij de gevolgde normeringsprocedure. *Tijdschrift voor Onderwijsresearch, 20*, 28-33.

Zwarts, M.A. & Janssens, F.J.G. (1994). *Meten van kerndoelen met PPON-toetsen*. De Meern: Inspectie van het Onderwijs.

Manuscript ontvangen 9-8-1994
Definitieve versie ontvangen 7-11-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, pp. 42-62

Classroom composition and individual achievement. Effects of classroom composition and teacher goals in Dutch elementary education[1]

Hans Luyten[a] and Anneke van der Hoeven-van Doornum[b]
[a] *University of Twente, Faculty of Educational Science and Technology**
[b] *ITS Nijmegen (Institute for applied social sciences)*

ABSTRACT

The present study deals with the consequences of classroom composition on individual achievement. The main research question is which effects are to be expected when students in the Dutch system of elementary education are grouped into more homogeneous classes. It seems likely that classroom homogeneity will increase in the near future, as parents of Dutch origin tend to send their children to schools where students of foreign origin constitute no more than a (small) minority of the entire school population. The investigations specifically focused on two possible effects of more homogeneous classrooms: changes with respect to the variation in student achievement and changes pertaining to the average level of student achievement. The effects of individual cognitive aptitudes, of the classroom average, of the classroom heterogeneity and their interaction effects were examined. It was also investigated to what extent the cognitive goals teachers set for their students can account for classroom composition effects on achievement. The analyses showed that an increase in classroom homogeneity may lead to a somewhat higher level of average achievement in Dutch elementary education, but also to a considerably larger variation. The achievements of the least talented students are likely to decrease. Although teacher goals were found to exert a considerable effect on individual achievement, they could not account for the observed classroom composition effects. The effect of the classroom average on individual achievement was found to be of similar size as the effect of socio-economic background.

1. INTRODUCTION

Most of the empirical research dealing with the effects of classroom heterogeneity in elementary education relates to the American educational system, where tracking students into separate classes according to their perceived cognitive aptitudes is a very common practice (Slavin, 1987). In the Netherlands students are hardly ever deliberately grouped into homogeneous classes. Most elementary schools are even too small to allow for such grouping practices. On average the elementary schools comprise only one class per grade. Because their size is rather small, the number of schools is quite large and it is not unusual to find several elementary schools within very close range, often in one and the same street and sometimes even in the same building.[2] Together with churches, pubs and bus connections elementary schools are the most widely available facilities in the Netherlands. For 90% of the population between the ages of four and twelve at least two elementary schools are available in their own residence (Blank et al., 1990; pp. 39-66).

Even though students are not deliberately grouped into homogeneous classes, many of them might end up in homogeneous classes, *because* the schools are so small and, as a result, mainly enrol students from a very restricted area. Especially the populations of urban schools are generally quite homogeneous with respect to the socio-economic and ethnic background of their

* Address: P.O. Box 217, 7500 AE Enschede.

students. Since the cognitive aptitudes that enhance success at school coincide with such background characteristics to a considerable extent, the student populations of urban schools may be quite homogeneous with respect to their students' aptitudes as well. Another important reason for this homogeneity, is the fact that parents of Dutch origin tend to send their children to schools where the students of foreign origin constitute no more than a (small) minority of the entire school population. As a result many students of foreign origin[3] end up in schools where Dutch students are largely absent (Rath, 1991; pp. 177-238).

The fact that in the Netherlands students in elementary schools are not deliberately grouped into homogeneous classes entails some important consequences with respect to the interpretation of the research outcomes to be reported. When students are deliberately tracked into homogeneous low-ability and high-ability classes, this is done to provide the most appropriate level and pace of instruction. In such cases it is no surprise that the teachers in high-ability classes set higher goals for their students than the teachers in the low-ability classes and it is therefore difficult to establish whether the less demanding instruction in low-ability classes is an appropriate response to the students' abilities or that it unnecessarily holds them back (Gamoran, 1992). In Dutch elementary education, however, the same curriculum is supposed to be taught across all classes. If Dutch teachers are found to set consistently lower goals in the low-ability classrooms this can be more straightforwardly interpreted as a response to the average ability level of the students.

2. PRESUMED AND OBSERVED EFFECTS OF CLASSROOM COMPOSITION

Grouping students into heterogeneous classes has frequently been advocated as a means to obtain an equitable distribution of student achievement. This is based on the often reported research finding that student achievement is positively (but moderately) affected by the classroom's average level of ability. Such effects have been reported for both elementary and secondary education in several educational systems, such as the American, Israeli, Dutch, English and Scottish (Beckerman & Good, 1981; Leiter, 1983; Rowan & Miracle, 1983; Dar & Resh, 1986; Kerckhoff, 1986; Willms, 1986; Mensen & Guldemond, 1987; Link & Mulligan, 1991; Resh & Dar, 1992; Reezigt, 1993). Grouping students into homogeneous classes is believed to reinforce already existing inequalities between students with respect to their cognitive aptitudes, because the less talented will end up in low-ability classes, which will affect their achievements negatively, while the more talented experience a positive effect of the high ability level in their classes. Slavin (1987), however, argues that assigning students to separate classes which are homogeneous with respect to their *general* cognitive abilities is less likely to have an effect on achievement than other forms of ability grouping. According to Slavin grouping plans can be expected to produce stronger effects when they reduce the heterogeneity with respect to the specific skill being taught, but classes will remain quite heterogeneous with respect to most subjects, when the same segregation criterion is applied for each subject. Teachers are then still faced with the problem of providing the appropriate level and pace of instruction for all the students in the classroom (Gamoran, 1992). The outcomes of the meta-analyses by Slavin (1987; 1990) dealing with the effects of ability grouping in elementary and secondary education support this view.

Another argument that has been advanced by proponents of heterogeneous grouping is that it might lead to a higher level of student achievement, because the effect of the average classroom ability is believed to be stronger for low-ability than for high-ability students (Oakes, 1985; Hallinan, 1987; De Vos, 1986; 1989). According to this line of reasoning the achievements of the most talented may be slightly impeded, but the achievements of the less talented are expected to rise more substantially. If all students would be grouped into heterogeneous classes instead of homogeneous ones, the talented students would end up in classes with an average ability level that is lower than their own individual level of ability and the reverse would be true for the low-ability students. If the effect of the average classroom ability is the same for the

talented and the less talented students, the overall effect of heterogeneous grouping would be zero. However, if the effect is stronger for the less talented students, then the achievement gain of the low-ability students would outweigh the loss of the high-ability students. Apart from the research outcomes presented by Dar & Resh (1986) and Resh & Dar (1992) which relate to secondary education in Israel there is not much empirical support for this view, which actually presumes an interaction effect of the individual and the classroom ability level on achievement. In figure 1 such an interaction effect is graphically displayed. Two lines are drawn: the upper line represents the effect of classroom composition for the talented students, while the lower line represents the effect for the less talented. In this artificial example the effect of classroom composition on achievement is much stronger for the less talented students.

It should be noted, that, even if such an interaction effect is actually at work, it does not always make sense to subtract the disadvantage of the talented students from the profit of the low-ability ones. Under certain circumstances more value must be attached to the achievements of specific groups and less to the achievements of the others. Consider the following two situations:

1. A large percentage of the students leaving elementary education appears to be virtually illiterate;
2. An increasing number of the students from pre-university education turns out not to be able to succeed in college.

If the number of illiterate students at the end of elementary schooling is unacceptably high, this would call for policy measures aiming to raise the reading achievements of the low-ability

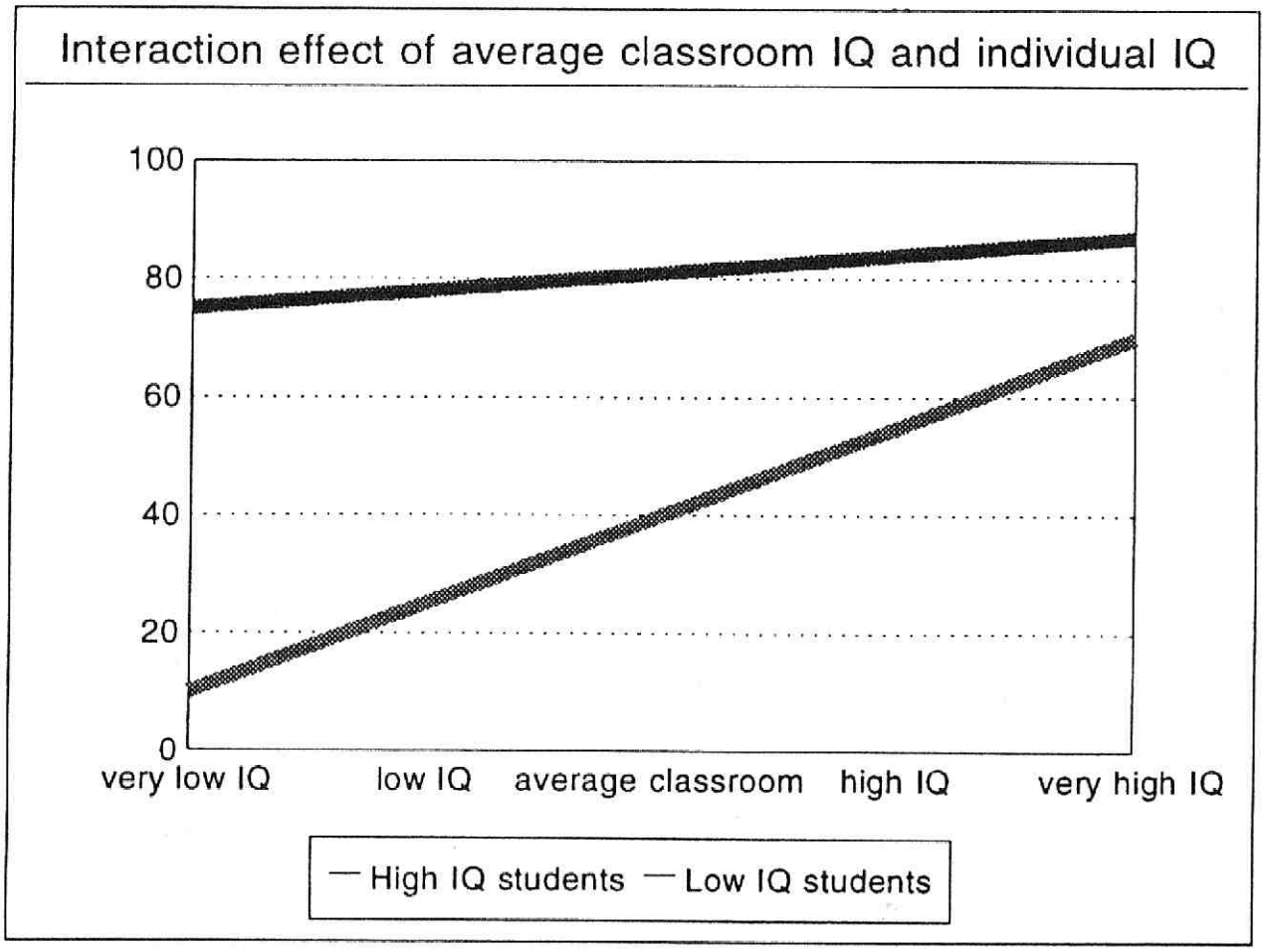


Fig. 1. Artificial example.

students even if this would lead to a considerable decline in the achievements of the more talented ones. In this case an approach resulting in a substantial reduction of the number of illiterate students would be needed, even if this would entail a decreased average achievement level. The second situation is more or less the reverse of the former. This time the achievements of the high-ability students are unacceptably low. Measures producing a rise in the achievements of the most talented students, but at the same time causing a downfall in average achievement may in this case be appropriate.

Another argument against tracking students into homogeneous classes that has frequently been raised is its contribution to the segregation of students from different social and ethnic backgrounds (e.g. Slavin, 1987; Gamoran, 1992).

Opponents of grouping students into heterogeneous classes maintain that teachers in homogeneous classes can more easily provide instruction at the most appropriate pace and level (Kulik & Kulik, 1982; 1984). According to this line of reasoning both high and low-ability students will suffer from heterogeneous grouping. The level and pace of instruction in heterogeneous classes will hardly provide any challenge for the high-ability students, while the low-ability ones are thought to be discouraged when they are confronted with highly talented classmates. Apart from two recent studies dealing with elementary education in the Netherlands (Maas, 1992; Reezigt, 1993) there is hardly any empirical evidence available that confirms the assertion that classroom heterogeneity affects student achievement negatively. Although the meta-analyses by Kulik & Kulik (1982; 1984) provide some support for the idea that ability grouping enhances student achievement in both elementary and secondary education, the outcomes do not allow for the conclusion that grouping students into homogeneous classes affects achievement positively. The Kuliks do not distinguish between several forms of ability grouping, such as assigning students to within-class groups for particular subjects, tracking them into separate classes or offering special programs for the gifted. In the meta-analyses by Slavin (1987; 1990) the effects of several types of ability grouping were separately investigated. Slavin concludes that the overall effect of tracking students into homogeneous classes on achievement is zero, both in elementary and secondary education. This conclusion is mainly based on American research. Positive effects of classroom heterogeneity on achievement have been reported for students in Israeli secondary education (Dar & Resh, 1986; Resh & Dar, 1992), while in Dutch secondary education classroom heterogeneity does not seem to affect the school careers of the students (De Vries, 1992).

Although the available evidence is not conclusive with respect to the direction of the heterogeneity effect, the findings hardly show any contradiction as far as the size of the effect is concerned. It is not clear whether the effect of classroom heterogeneity on achievement is positive or negative, but is does not seem very strong.

Some other possible effects of classroom composition that are not frequently mentioned in the debate on the advantages and disadvantages of classroom heterogeneity require some attention as well. First of all, the interaction of individual ability and classroom heterogeneity. Grouping students into heterogeneous classes is sometimes believed to result in a higher level of general achievement, because low-ability students are expected to be more sensitive to the average level of ability in their classroom than their more talented classmates. However, if low-ability students are more sensitive to the classroom ability level, they may also be more sensitive to other classroom characteristics, such as the possibly negative effect of classroom heterogeneity. This possibility should be taken into account in an analysis that deals with the effects of grouping students into heterogeneous classes, for it may neutralize the positive effect of the average classroom ability on the achievements of the low-ability students. Secondly, the interaction of the average classroom ability and heterogeneity. The effect of the average ability level may be stronger in homogeneous classes, because in those classes teachers can provide instruction at the same level and pace for the entire classroom. The equalizing effect that is believed to result from heterogenous grouping may then be rather weak. If all students are grouped into maximally heterogeneous classes and the average ability level is identical across all classes, the impact of the average classroom ability on individual achievement may be minimal. Thirdly,

there may be a three-way interaction effect of individual ability, classroom ability and classroom heterogeneity, implying that the effect of classroom ability and heterogeneity may be different for low-ability and high-ability students.

3. CLASSROOM COMPOSITION EFFECTS AND RATIONAL CHOICE THEORY

The explanations that have been proposed for the effects of classroom composition on individual achievement can be grouped into two main categories. The first one refers to the idea of differential instruction (e.g. Beckerman & Good, 1981; Rowan & Miracle, 1983; Gamoran, 1986; 1992; Slavin, 1987; 1990) implying that high-ability classes receive a more favourable instruction, because teacher behaviour varies depending on the classroom composition or because the more competent teachers are disproportionately assigned to high-ability classrooms. In the other category of explanations the role of interaction and competition among classmates is emphasized (e.g. Erbring & Young, 1979; Mensen & Guldemond, 1987; Duke, 1993). Students in high-ability classes are believed to be stimulated by the achievements of their peers, whereas students in low-ability classes lack such incentives.

In this paper we will focus on the differential instruction hypothesis in order to explain the effects of classroom composition on individual achievement. De Vos (1986; 1989) has pointed out that the idea of differential instruction can be incorporated into rational choice theory quite easily. Differential instruction can be conceived as resulting from purposeful behaviour of both teachers and students. De Vos has summarized his argument as follows:

> "students' efforts are an important determinant of their achievements, ... the teacher's evaluative behavior affects the amount of efforts students are willing to expend on school work, and ... students' achievements affect the teacher's evaluative behavior." (De Vos, 1989, p. 223)

Both teachers and students are assumed to maximize their "subjective expected utility", which results in striving for social approval and physical well-being. Students can obtain social approval from the teacher through achievement. Achieving, however, requires effort. The efforts, which lead to a decrease in physical well-being, represent the costs of achievement. It goes without saying that the costs of achieving are relatively high for low-ability students and low for high-ability students. According to De Vos students will keep raising their achievements up to the point that the marginal utility of achieving equals its marginal cost. Going beyond this point would be irrational, because raising one's achievements even higher would entail extra costs that are no longer offset by an extra amount of social approval. The amount of utility a student obtains as a result from achievement is largely dependent on the way the teacher responds to the efforts of the students.

Teachers are believed to obtain social approval through the achievements of their students. Like the students, they are faced with a problem of optimization, since raising the student achievements requires effort from the teachers as well. The intellectual capacities of the students also determine the costs for the teachers to a large extent, for it will be easier to obtain high achievements in high-ability classes. It is assumed that teachers behave basically in the same way as the students. Both teachers and students are believed to raise their efforts up to the point that the marginal utility equals the marginal cost. As a result teachers will require higher achievements, i.e. apply a more demanding standard, in classes with a high average ability level and lower achievements in low-ability classes. Students in high-ability classes, also the less talented, may therefore be expected to reach a higher level of individual achievement, because they are confronted with a more demanding standard.

It should be noted that De Vos strongly emphasizes the effect of standards on the classroom as a whole, although he acknowledges that teachers do not apply a uniform standard to all their students. This would in most cases even be irrational, because the standard would then be too

high for some students in the classroom, and too low for others. Students who are faced with standards that are set too high for them, will be strongly discouraged to expend any effort at all, while talented students who are confronted with a very low standard, will hardly be encouraged to raise their achievements, even though they are able to do so. The crucial point, however, is that students of equal ability are believed to be confronted with different standards depending on the average classroom ability level. Teachers in high-ability classes can be expected to set higher individual goals for each student and the students in those classes will be faced with a higher general classroom standard.

De Vos also contends that classroom composition effects are stronger for low-ability students than for the more talented ones and that this can also be conceived as resulting from purposeful behaviour of the teacher, if one is willing to accept the additional assumption that teachers in low-ability classes tend to pay relatively much attention to their best students in order to "prove" that the low achievements in their class are not caused by poor instruction. This implies that teachers in low-ability classes can be expected to set relatively high standards for their most talented students.

4. RESEARCH QUESTIONS

Our main research question refers to the effects of classroom composition on achievement. We start our analyses assessing the classroom homogeneity with respect to the socio-economic background and general intelligence of the students. In addition we examine to what extent classroom composition affects the standards the teachers employ and to what extent these standards can account for the effects of classroom composition. The analyses specifically address the following questions:

1. To what extent are the classrooms in Dutch elementary education homogeneous with respect to the socio-economic and intellectual backgrounds of the students? This question will be dealt with in section 6.1. A substantial amount of classroom homogeneity is expected, especially with respect to the students' socio-economic backgrounds, because most schools enrol students from a rather restricted area.
2. In how far do teachers apply a uniform standard for the entire classroom? We will address this question in section 6.2. In the explanation of classroom composition effects proposed by De Vos (1986; 1989) the effect of standards on the classroom as a whole is strongly emphasized. On the other hand, it is not plausible that teachers will apply a perfectly uniform standard for all their students.
3. Does classroom composition affect the goals teachers set for each individual student? Do teachers set higher goals in high-ability classes independently of the students' individual abilities? According to De Vos teachers set higher goals for their students in high-ability classes. Furthermore, teachers in low-ability classes are believed to set particularly high goals for their most talented students and relatively low goals for the less talented. Section 6.2 presents an empirical assessment of the hypothesized relations between classroom composition and teacher goals.
4. To what extent do the average classroom ability level and classroom heterogeneity affect the individual achievements of the students independently of their individual ability and their socio-economic family background? We will deal with this question in section 6.3. A positive effect of the average classroom ability level is expected, which would imply that grouping students into homogeneous classrooms reinforces the already existing inequalities with respect to their intellectual capacities. A negative effect of classroom heterogeneity, however, would imply that more equality with respect to achievement entails a somewhat lower level of general student achievement.
5. Are the effects of classroom composition different for high and low-ability students? Is the effect of the average classroom ability stronger in homogeneous or in heterogeneous classes?

In other words: are there any significant interaction effects of individual ability, classroom ability and heterogeneity? These issues are addressed in section 6.3. Investigating the strength of these interaction effects is required for an adequate understanding of the effects of classroom composition. One of the arguments advanced by the proponents of heterogenous grouping is the assertion that heterogeneous grouping will raise the general level of student achievement, because the positive effect of the average classroom IQ is believed to be particularly strong for the less talented students. On the other hand, one should acknowledge that this effect may be counterbalanced by a negative heterogeneity effect which mainly affects the achievements of the low-ability students. It should also be checked whether the effect of the classroom ability level remains equally strong when classes become more heterogeneous.

6. To what extent can the goals set by the teachers account for the effects of classroom composition? How strong are the classroom composition effects after controlling for teacher standards? How strong is the empirical basis for the conception of classroom composition effects as resulting from purposeful teacher behaviour? These questions are discussed in section 6.3.
7. Which changes are to be expected if the students were grouped into perfectly heterogeneous or perfectly homogeneous classes? How will it affect the variation in student achievement? What will be the consequences for high and low-ability students. These questions will be dealt with in section 6.4.

The next section presents a description of the datasets that were used in the analyses. The outcomes are reported in section 6.

5. DATA AND METHOD OF ANALYSIS

Two different datasets were analyzed to investigate the effects of classroom composition. The first dataset, which contains information about a national sample of 212 elementary schools, was used to assess the effects of classroom composition, whereas the second dataset was primarily used to test whether teacher goals can account for the effects of classroom composition. The first dataset has been used by several researchers in various studies (e.g. Brandsma, 1993; Knuver, 1993; Reezigt, 1993). The present study focuses on the students who were in their final year of elementary education. The data relate to 3993 students from 221 classes and were collected in 1987 and 1988. To assess the impact of classroom composition on individual achievement information about the following three variables was needed:

- achievement
- intelligence
- socio-economic status (SES)

Classroom characteristics, such as means and standard deviations, could be computed on the basis of this information. To measure achievement a test made up of 88 items relating to language and mathematics was used.[4] These items were derived from standardized tests ("CITO-tests"), which are administered in most schools a few months before the end of the final year in elementary education. The CITO-tests can be considered to cover the basic curriculum in Dutch elementary education quite adequately (Blok, 1992). Intelligence was measured one year earlier, in 1987, by means of an IQ-test (the "ISI-test") measuring both spatial and verbal intelligence. Information about the socio-economic background of the students, which served as a control variable in our analyses, was obtained from the teachers. The SES-index is based on four indicators: the education of a student's father and mother and their profession.

The second dataset contains information about a sample of 51 schools from four regions in the East and South of the Netherlands (Twente, Arnhem, Tilburg and Zuid-Limburg). The data were originally collected at several points in time between may 1986 and june 1988 by Van der Hoeven-Van Doornum (1990) for a study into the indirect effect of socio-economic

status via teacher expectations and aspiration levels on school careers in elementary education. Students were followed during their last two years of elementary education. In the present study we confine ourselves to the effects of classroom composition in the final year of elementary education. The dataset was obtained through disproportionate sampling (Van der Hoeven-Van Doornum, 1990; pp. 48-58). Two stratification criteria were applied: the composition of the schools' student populations with respect to socio-economic and ethnic background and the academic output of the schools in 1985.[5] Schools were grouped into three categories with respect to the first criterion (high, medium and low-input schools). On the basis of the second criterion the schools in each category were classified as high, low or medium-output schools. From each category an equal number of high and low-output schools was included in the designed sample. The medium-output schools were not included in the sample.[6] The schools in the sample did not differ significantly from the total population of schools in the four regions with respect to student enrolment, denomination (public, catholic, protestant or other), degree of urbanization or the socio-economic and ethnic background of their students (Van der Hoeven-Van Doornum, 1990; pp. 59-60). The students in the sampled schools scored slightly below the national average on the test score that served as the main criterion variable in our analyses – 533.3 versus 535.7 – , while the standard deviation in the sample was somewhat higher than in the national population – 11.25 versus 9.75 (Van der Hoeven-Van Doornum, 1990; p. 94). The analyzed dataset contained information about 698 students from 57 classes. Our analyses required information about only four student level variables. Classroom characteristics were computed on the basis of these four variables, which are:

- achievement
- intelligence
- socio-economic status (SES)
- cognitive goals set by the teacher

Achievement was measured by means of a standardized test (the "final CITO-test") consisting of 180 multiple choice items dealing with language, mathematics and information processing (i.e. reading tables, maps and graphs). This test is administered in the majority of Dutch elementary schools shortly before the end of the final year, usually in february. Intelligence was assessed by means of the same IQ-test (the "ISI-test") as the one in the national sample. This test was administered in september 1987. Socio-economic background was measured by an index constructed on the basis of the education and profession of the students' parents. The information was obtained directly from the parents. The cognitive goals were measured by a nine item index, which serves as an operationalization of their standards. In december, three months after the beginning of the school year, the teachers were interviewed about the cognitive goals they had set for each of their students. The questions were formulated as follows:[7]

Please indicate your approach for this student.
- Set modest standards*
- Emphasize practical work rather than language and mathematics development*
- Teach only topics the student will need later on*
- Let the student work with material from lower grades*
- Teach extensive parsing skills
- Give the student the opportunity to switch to easier subject matter*
- Present enrichment material
- Let the student work ahead
- Demand correct spelling

These nine questions were Likert-type items with five response categories. The internal consistency of the scale (Cronbach's α) equalled .86. The items marked with an asterisk (*) were negatively scored.

It should be noted that in the present study classroom homogeneity is conceived as a sociological phenomenon rather than an educational one. In Dutch elementary education students are hardly ever deliberately grouped into homogeneous classes. As far as students do end up in homogeneous classes this is the result of sociological factors. Schools generally enrol students from restricted areas, which are often quite homogeneous with respect to the socio-economic and ethnic backgrounds of the residents and Dutch parents tend to send their children to schools where the students of foreign origin constitute no more than a (small) minority of the entire school population, so that many students of foreign origin end up in "black" schools. We expect that such socio-economic and ethnic homogeneity produces a substantial homogeneity with respect to the cognitive aptitudes of the students as well. The present study aims to establish in how far the classroom composition characteristics with respect to the general cognitive aptitudes of the students affect their individual achievements. It was decided to investigate the effect of the (individual and classroom average) IQ-scores rather than the effect of prior achievement, because IQ-scores are more stable in time and less curriculum-dependent than test-scores which relate to more specific knowledge and skills. The analyses aim to assess the effect of classroom composition on the general achievement scores of the students and do not differentiate between language and mathematics achievement.

Several multilevel analyses were conducted to answer the research questions formulated in section 4. In order to obtain an impression of the effects to be expected if the students were grouped into more heterogeneous or homogeneous classes, some simple simulations were conducted based on the empirical findings with respect to the classroom composition effects.

6. RESULTS

The multilevel analyses were conducted using the ML3-software (Prosser et al.; 1991). The findings are presented in the next three sections. The outcomes with respect to the simulations are presented in section 6.4.

6.1. Classroom homogeneity with respect to IQ and SES

The homogeneity of the classrooms in the investigated sample with respect to the socio-economic background and the individual IQ-scores of the students was assessed by means of a multilevel analysis. For both variables the total amount of variance was partitioned into student level and classroom level variance. The amount of classroom level variance expresses how much of the total variance can be attributed to differences between the classroom means. The student level variance expresses how much of the total variance is attributable to differences between students within classes. The results are listed in table 1, which shows that, although most of the variance is situated at the student level, there are still substantial differences between classes with respect to the socio-economic backgrounds of their students. The differences between classrooms with respect to the IQ-scores were found to be considerably smaller, especially in the national sample. No more than 11.3% of the total variance in IQ-scores appeared to be situated at the classroom level, whereas about a quarter of all variance with respect to the SES-scores turned out to be attributable to differences between classes. This implies that the

Table 1. Student and classroom level variance with respect to IQ and SES.

	National Sample		Regional Sample	
	SES	IQ	SES	IQ
Student level variance	74.9%	88.7%	73.2%	81.9
(standard errors)	(1.7)	(1.6)	(4.1)	(4.6)
Classroom level variance	25.1%	11.3%	26.8%	18.1%
(standard errors)	(2.9)	(2.0)	(6.5)	(4.9)

classes in Dutch elementary education cannot be made much more heterogeneous than they presently are with respect to the cognitive aptitudes of the students.

The appendix provides additional information about the frequency distributions of these and other variables that were analyzed in the present study. The correlation between socio-economic status and IQ was found not to be very high at the individual level in either sample. The relation between both variables appeared to be stronger at the classroom level. The student level correlation in the national sample is .33 (pearson r), and .37 in the regional sample, while the correlation between the classroom aggregates equals .49 in the national sample and .57 in the regional sample.

In several studies which deal with the effects of classroom composition the researchers have been faced with the problem that the average classroom ability level and the classroom heterogeneity were very strongly correlated (Guldemond et al., 1987, pp. 61-75). In the present study the standard deviation of the IQ-scores in the classroom served as a heterogeneity measure. The correlations between the average classroom IQ and the standard deviation did not lead to any problems of multicollinearity in either sample. The correlation between both variables equalled -.13 in the national sample (which is not significant for $\alpha < .05$) and -.44 in the regional sample.

6.2. Effects of classroom composition on teacher goals

This section deals with the question to what extent teachers set uniform goals for all the students in their classes and to what extent classroom composition variables can account for the goals teachers set for the individual students. The second dataset, the one relating to the regional sample, was used to address these issues. The total variance in teacher goals was partitioned into student level and classroom level variance. Next the effects of individual student characteristics, classroom characteristics and their interaction effects on the teacher goals were examined. The results of these analyses are presented in table 2. Before conducting the analyses all variables were transformed into z-scores, so that the reported effects can be interpreted as standardized regression coefficients. The figures in between brackets denote the standard errors. The interaction terms were computed by multiplying the z-score transformations. The effects of these interaction terms can be interpreted as standardized regression coefficients as well (Jaccard et al., 1990).

Model 0 shows the partitioning into student and classroom level variance. The percentages of variance at the student and classroom level are similar to the ones found with respect to socio-economic background and IQ-scores. Most of the variance in teacher goals is situated within the classrooms, which implies that teachers set quite different goals for the students in their classes. Although the amount of classroom level variance is rather modest as compared to the student level variance, the outcomes do reveal substantial differences between classes with respect to teacher goals. Model 1 presents the effects of several independent variables and their interactions on the teacher goals. The regression coefficients that are significant for $\alpha < .05$ in a two-tailed t-test are printed in bold face. Most coefficients, however, are very small and not statistically significant. Individual IQ reveals the strongest effect on the goals set by the teachers. Teachers also set higher goals for students from a more favourable socio-economic background. Apart from the interaction effect of classroom heterogeneity (measured by the classroom standard deviation for IQ-scores) and individual IQ, no significant effects were found for the classroom composition characteristics. The interaction effect implies that in the more heterogeneous classes teachers set particularly high goals for the most talented students, and particularly low goals for the less talented.

The expectation that teachers set higher goals in high-ability classes was not confirmed. Teachers do set higher goals for the more talented students, but this is not related to the average level of ability in the classroom. Also, the expectation that teachers in low-ability classes set relatively high goals for their most talented students, which would imply a negative interaction effect of individual and classroom ability, cannot be considered to be corroborated. However, the analyses did reveal substantial differences between classrooms with respect to the goals set by the teachers. The average classroom goals may have an effect on student achievement.

Table 2. Effects of individual and classroom characteristics on teacher goals; Regional Sample.

	Model 0	Model 1
Standardized regression coefficients (standard errors)		
1. Individual characteristics		
IQ-score	–	**.511** (.035)
Socio-economic status	–	**.092** (.033)
2. Classroom characteristics		
Mean IQ	–	-.040 (.068)
Standard deviation IQ	–	-.004 (.062)
3. Interaction effects		
Individual IQ * mean IQ	–	-.020 (.033)
Individual IQ * std. dev. IQ	–	**.076** (.031)
Mean IQ * std. dev. IQ	–	-.053 (.045)
Ind. IQ * mean IQ * std. dev. IQ	–	-.003 (.025)
Grand Mean (standard errors)	-.073 (.070)	-.053 (.066)
	VARIANCE	VARIANCE EXPLAINED
Student level	79.1%	37.7%
Classroom level	20.9%	21.9%
Total	100.0%	34.3%

6.3. Effects of classroom composition and teacher goals on individual achievement

In this section it will be discussed to what extent classroom composition affects student achievement and whether teacher goals can account for such effects. Both the national and the regional sample are examined. The information provided by the national sample allows for a reliable estimation of the impact of classroom composition in Dutch elementary education. An analysis of the regional sample will show in how far the effects of classroom composition can be attributed to the goals set by the teachers.

The zero models in table 3 show that the amount of classroom level variance with respect to student achievement is substantial in both samples. The differences between classrooms in this respect are considerably larger than the classroom differences regarding IQ-scores and teacher goals. They are also somewhat larger than the differences with respect to socio-economic background. The classroom level variance probably provides a close approximation of the school level variance, because only a few schools in both samples comprise more than one class in grade 8. The classroom level variance thus largely coincides with the school level variance. In most studies dealing with school effectiveness in Dutch elementary education the amount of school level variance, however, has been reported to be considerably smaller (Blok, 1992). It is quite likely that the differences between schools are somewhat overestimated in the regional sample, because schools with remarkably high and low outputs in former years are overrepresented to some extent (see section 5). It seems less likely, though, that the relatively large amount of school level variance that was found in the national sample is due to the sampling design. It should be noted that many reports which deal with the differences between Dutch elementary schools are not based on representative samples (Blok, 1992). In some cases the sample only relates to a certain region (Meijnen, 1984; Bosker & Hofman, 1987), in some cases schools with many disadvantaged students are overrepresented (Van der Werf & Weide, 1991; Van der Werf et al., 1991; Weide, 1993; Van der Velden, 1993; Jungbluth, 1993) and in other cases there is no information about the representativeness of the samples (Brandsma & Knuver, 1988; Van de Grift & Akkermans, 1991). Apart from the national sample that was analyzed in

the present study the sample analyzed by Blok (1992) may be the only one that can be considered to produce a representative picture of the differences between schools. This sample contains information about student achievement in 179 schools for five consecutive years (1987-1991). Blok reports a declining trend in time with respect to the amount of school level variance. For 1988, the year when the data for the national sample in the present study were collected, he reports a school level variance of 17% for language and 22% for mathematics. These outcomes still deviate to some extent from the findings in the present study, but they also suggest that the relatively small amounts of school level variance reported in previous studies underestimate the true differences between schools.

For both samples model 1 presents the effects of individual characteristics and classroom composition on student achievement. Again, the reported effects can be interpreted as standardized regression coefficients and the ones which are significant for $\alpha < .05$ in a two-tailed t-test are printed in bold face. The strongest effect was, not surprisingly, presented by individual intelligence. Socio-economic background was found to affect achievement as well. A moderate, but statistically significant effect of the average classroom ability level could be detected in both samples. The national sample also revealed a significant effect of classroom heterogeneity and a significant interaction effect of individual and classroom IQ. Classroom heterogeneity was found to exert a modestly negative effect on individual achievement. The negative sign of the interaction effect implies that the effect of the average classroom IQ is stronger for low-ability students than for the more talented ones. This effect, however, is even smaller than the effect of classroom heterogeneity. Similar interaction effects and heterogeneity effects were found in the regional sample as well, but in that case they were not statistically significant. This is probably due to the limited size of this sample which renders the statistical power of the analyses rather low if one wants to detect effects of a modest size.[8] In general the findings from both samples are fairly consistent. It is difficult, however, to establish which factors may account for the observed inconsistencies. The fact that the samples were obtained through different sampling designs may be the most important explanation for the divergencies. The national sample must be considered to provide a more representative picture, because it relates to the entire system of Dutch elementary education and because in the regional sample schools with high and low output are overrepresented. It should also be noted that two variables, socio-economic status and student achievement, were not measured in exactly the same way in both samples. The achievement scores in the national sample relate only to language and mathematics, whereas the achievement scores in the regional sample are based on items pertaining to language, mathematics and information processing. In the regional sample the information with respect to the students' socio-economic background was obtained directly from the parents, but in the national sample the teachers were asked to provide this information. It can of course not be precluded that some of the inconsistencies reflect certain genuine regional peculiarities in the East and South of the Netherlands, but this does not seem very likely.

The fact that the other two-way interaction effects were found to be very small and non-significant in the national sample is also worth mentioning. Our concern that the low-ability students might be more strongly influenced by a negative heterogeneity effect than the more talented ones (see section 2) turned out not to be justified. The same is true for our worries that the impact of the classroom average might decrease in heterogeneous classrooms. The regional sample even displayed (non-significant) effects in the opposite direction. The three-way interaction of individual IQ with classroom IQ and classroom heterogeneity also appeared to be insignificant in both samples.

When controlling for the cognitive goals set by the teacher in the regional sample, some remarkable changes emerge. Model 2 shows a considerable, positive effect of the goals set by the teacher at the individual level, but the classroom average of the teacher goals was found to affect achievement negatively. This effect, however, is not significant for $\alpha < .05$. This finding is consistent with the outcomes presented by Van der Velden (1993) who investigated the effect of a similar variable on language and mathematics achievement. In his analyses the teacher goals are related to the classroom as a whole and not to individual students. This variable was

Table 3. Classroom composition, teacher goals and individual achievement.

	National Sample		Regional Sample		
	Model 0	Model 1	Model 0	Model 1	Model 2
Standardized regression coefficients (standard errors)					
1. Individual characteristics					
IQ-score	–	**.577** (.012)	–	**.658** (.026)	**.468** (.026)
Socio-economic status	–	**.167** (.012)	–	**.128** (.024)	**.095** (.021)
Teacher goals	–	–	–	–	**.371** (.025)
2. Classroom characteristics					
Mean IQ	–	**.107** (.024)	–	**.128** (.050)	**.188** (.055)
Standard deviation IQ	–	**-.070** (.025)	–	-.048 (.046)	-.043 (.043)
Classroom mean teacher goals	–	–	–	–	-.097 (.055)
3. Interaction effects					
Individual IQ * mean IQ	–	**-.039** (.010)	–	-.034 (.025)	-.026 (.022)
Individual IQ * std. dev. IQ	–	-.013 (.011)	–	.024 (.023)	-.004 (.020)
Mean IQ * std. dev. IQ	–	-.001 (.018)	–	.058 (.034)	**.076** (.032)
Ind. IQ * mean IQ * std. dev. IQ	–	.010 (.008)	–	.030 (.019)	.031 (.025)
Grand Mean (standard error)	-.057 (.039)	-.000 (.026)	-.094 (.087)	-.005 (.014)	-.012 (.048)
	VARIANCE	VAR. EXPL.	VARIANCE	VARIANCE	EXPLAINED
Student level	71.3%	49.4%	64.9%	61.0%	70.8%
Classroom level	28.7%	60.1%	35.1%	75.9%	76.1%
Total	100.0%	52.5%	100.0%	66.2%	72.7%

not found to affect achievement. Another variable, however, which relates to the minimum goals the teacher sets for his/her classroom did produce a significant effect. We will return to this question in the discussion section.

The outcomes suggest that at the individual level the teacher goals function as an intermediate variable between individual IQ and socio-economic background on the one hand and achievement on the other. The effects of both variables on achievement in the regional sample are considerably lower in model 2 than in model 1. The effects of individual IQ and socio-economic background in model 1 can be considered as the total effects on achievement, whereas model 2 presents the direct effects, i.e. the effects that are independent of the teacher goals. In the previous section the teacher goals were shown to be dependent on IQ and SES. The difference between the regression coefficients for the interaction effect of classroom heterogeneity with individual IQ presents another indication for the supposition that the teacher goals intermediate between individual ability and achievement. In section 6.2 a significant interaction effect on teacher goals of individual ability with heterogeneity was reported. Model 1 in table 3 presents an interaction effect in the same direction on achievement for the regional sample, although it is not statistically significant. When controlling for teacher goals, however, the interaction effect is virtually equal to zero. The effects of the average classroom ability level and its interaction with classroom heterogeneity appear to increase when controlling for teacher goals. In model 2 this interaction effect is even significant for $\alpha < .05$, which would imply that the effect of the average classroom IQ is *stronger* in heterogeneous classroom instead of weaker. Our worries that the effect of the average classroom ability level might be rather weak in heterogenous classes were therefore not confirmed. The results presented in model 2 clearly indicate that the goals set by the teachers, as operationalized in the present study, cannot account for the classroom composition effects that were found.

Table 4. Models for simulating student achievement (based on the national sample).

	Model 1	Model 2	Means and standard deviations	
Regression coefficients			**Mean**	**Std. Dev.**
Individual IQ-score	1.679	1.647	11.494	1.803
Socio-economic status	0.133	0.129	0.000	6.695
Classroom mean IQ	–	0.821	11.494	0.714
Standard deviation IQ	–	-1.020	1.673	0.323
Interaction Ind. IQ * mean IQ	–	-0.136	0.509	1.601
			Achievement score	
Grand Mean	1.119	-6.114	20.562	5.205
	VARIANCE	EXPLAINED		
Student level	49.3%	49.4%		
Classroom level	50.7%	60.0%		
Total	49.7%	52.4%		

6.4. Simulated effects of classroom composition

In order to obtain an impression of the effects to be expected if the students were grouped into more heterogeneous or more homogeneous classes some simple simulations were conducted based on the empirical findings in the national sample. The general model that was used to "predict" the effects is shown in table 4. The simulations are based on model 2. Model 1 is included in the table to show how much of the variance in student achievement can be explained by the classroom composition variables. The extra amount of variance explained by model 2 at the classroom level is substantial (9.3%), but its contribution at the individual level is very small (0.1%). The increase across both levels is therefore quite modest as well (2.7%). All regression coefficients in the table are significant for $\alpha < .01$.

Table 5. shows the outcomes of the simulations. Three situations were considered:

A. The present situation
B. A situation in which students were grouped into completely heterogeneous classes
C. A situation in which students were grouped into completely homogeneous classes.

ad A: First of all, the achievement scores that would be expected in the present situation on the basis of our model were computed. In this way a "yardstick" was obtained to evaluate the outcomes in the other two situations. The expected scores were computed by means of the following equation:

$$\mathbf{CITO_{exp}} = -6.114 + 1.647 * \mathbf{IQ_{ind.}} + .129 * \mathbf{SES} + .821 * \mathbf{IQ_{class}} - 1.020 * \mathbf{SD_{IQ}} - .136 * \mathbf{IQ^c_{ind.}} * \mathbf{IQ^c_{class}}$$

Where:

$\mathbf{CITO_{exp}}$ = The expected achievement score on the CITO-test
$\mathbf{IQ_{ind.}}$ = The individual IQ-score
$\mathbf{SES}$ = The socio-economic background
$\mathbf{IQ_{class}}$ = The classroom mean IQ-score
$\mathbf{SD_{IQ}}$ = The standard deviation of the individual IQ-scores in the classroom
$\mathbf{IQ^c_{ind.}}$ = The individual IQ-score centred around the overall mean IQ-score (11.494)
$\mathbf{IQ^c_{class}}$ = The classroom mean IQ-score centred around the overall mean IQ-score (11.494)

The product of the last two variables makes up the interaction term that relates to the interaction effect of individual and classroom IQ. An appropriate procedure to assess a two-way interaction effect in multiple regression models is to centre both variables around their respective means before computing a multiplicative term. In this way the problem that the correlations between the interaction term and the variables $\mathbf{IQ_{ind.}}$ and $\mathbf{IQ_{class}}$ would be too large (the problem of multicollinearity) can easily be avoided (Cronbach, 1987, Jaccard et al., 1990).[9]

Table 5 shows the average and the standard deviation of the expected scores. Note that the standard deviation of these scores is considerably lower than the actual standard deviation. This is because our model accounts for only 52.4% of the total variance in the achievement scores. The table also shows the expected scores for students of different ability. For this purpose the students were grouped into four categories of equal size.

ad B: In this case each class would be a perfectly representative sample of the entire population of students. The average IQ and the standard deviation of IQ-scores in each class would equal the average score and standard deviation in the whole population. The value for $\mathbf{IQ_{class}}$ is thus assumed to be 11.494 in all classes. Since the regression effect for $\mathbf{IQ_{class}}$ is .821, the effect of this variable should equal 9.437 (.821*11.494) in all classes. The value for $\mathbf{SD_{IQ}}$ is assumed to equal 1.803 in all classes and the regression coefficient for this variable is -1.020. The expected effect of classroom heterogeneity is thus estimated to be -1.839 (-1.020*1.803) in each and every class. The interaction term is equal to zero in this situation, because the value for $\mathbf{IQ^c_{class}}$ is zero in every class. Summing the values for the intercept and the effects of $\mathbf{IQ_{class}}$ and $\mathbf{SD_{IQ}}$ yields 1.484 (-6.114 + 9.437 – 1.839). The equation to compute the expected scores thus becomes:

$$\mathbf{CITO_{exp}} = 1.484 + 1.647*\mathbf{IQ_{ind.}} + .129*\mathbf{SES}$$

ad C: In this hypothetical situation the students are assumed to be grouped into classrooms with zero heterogeneity. This implies that the negative effect of $\mathbf{SD_{IQ}}$ is totally absent and that for each student the average classroom IQ equals his or her individual IQ. The variable $\mathbf{IQ_{ind.}}$ now actually occurs two times in the equation, because $\mathbf{IQ_{ind.}}$ and $\mathbf{IQ_{class}}$ are assumed to be identical. This can of course be simplified by computing a single term ($2.468*\mathbf{IQ_{ind.}}$ instead of $1.647*\mathbf{IQ_{ind.}} + .821*\mathbf{IQ_{ind.}}$). Since in this situation $\mathbf{IQ_{ind.}} = \mathbf{IQ_{class}}$, the interaction term changes into a quadratic term. The equation now becomes:

$$\mathbf{CITO_{exp}} = -6.114 + 2.486*\mathbf{IQ_{ind.}} + .129*\mathbf{SES} - .136*(\mathbf{IQ^c_{ind.}})^2$$

Table 5 shows that extremely homogeneous grouping would result in much larger deviations from the present situation than extremely heterogeneous grouping. This is not surprising, because in the present situation the students are already grouped into fairly heterogeneous classrooms

Table 5. Simulated achievement scores in different grouping conditions.

	Present Situation (expected scores)	Heterogeneous Classrooms	Homogeneous Classrooms
High IQ students	25.0	24.7	27.4
Moderately high IQ	21.8	21.6	23.7
Moderately low IQ	19.4	19.3	20.3
Low IQ students	15.5	16.0	14.8
General Average	20.5	20.4	21.6
Standard Deviation	3.72	3.35	4.84

(see table 1). Homogeneous grouping may be expected to result in a somewhat higher level of average achievement, but also in considerably more variance. The achievements of the least talented would decrease, but the achievements of the others would increase, especially those of the brightest 50%. Grouping students into extremely heterogeneous classes would result in a slightly lower amount of variance and in somewhat higher achievement of the least talented students.

The figures in table 5 should be considered primarily as an illustration of the trends that were detected in the national sample rather than definite predictions of student achievement under certain conditions. It should be noted that the model on which the simulations are based, leaves nearly 50% of the total variance in student achievement unexplained. Moreover, the simulations are based on the assumption that the trends which were found in the present situation can be extrapolated to conditions of extreme homogeneity or heterogeneity of classrooms. Especially the situation of perfectly homogeneous classes deviates strongly form the prevailing circumstances.

7. DISCUSSION

The main purpose of the present study was to assess the effects of classroom composition on individual achievement in Dutch elementary education. It was specifically investigated what will be the consequences if students become grouped into more homogeneous classes. The research outcomes indicate that classroom homogeneity tends to reinforce the inequalities between students with respect to their cognitive aptitudes. Increasing classroom homogeneity may result in higher achievements for most students, especially the most talented, but it must be expected to impede the achievements of the least talented. Our concerns that the effect of the average classroom IQ on individual achievement might be very weak in heterogeneous classes or that the negative effect of classroom heterogeneity is particularly strong for the low-ability students were not confirmed. Although the classrooms in both samples appeared fairly homogeneous with respect to the socio-economic backgrounds of their students, the classes turned out to be surprisingly heterogeneous with respect to the cognitive aptitudes of the students.

Classroom homogeneity, however, must be expected to increase under the present circumstances. Dutch parents generally seem to prefer schools with no more than a small minority of foreign students for their children, because they suspect that a large number of foreign classmates affects the achievements of the Dutch students negatively. As a result many students of foreign origin end up in classes which are almost entirely made up of other foreign students. Without government intervention this development can be expected to result in a considerably increased homogeneity of the classes, which will coincide with a segregation along ethnic lines. The present government policy, which is aimed at increasing the size of the elementary schools, may provide some opportunities to counterbalance this trend. The larger the schools, the wider the area from which they enrol their students. The student populations of the schools may thus become more heterogenous as their size increases, although this does not guarantee that the classrooms will become more heterogeneous. Presently most schools are too small to allow for the grouping of students into homogeneous classes. When school size increases, some schools may choose to track students into high and low-ability classes. It seems dubious that the creation of larger schools can on its own turn the tide of increasing classroom homogeneity. More heterogeneous classes and a more equal distribution of student achievement can only be obtained through more specific measures. It should also be noted that, although the classrooms in the national sample were found to be quite heterogeneous with respect to the IQ-scores of their students, this does not preclude that in some regions, especially the large cities, classes may still be very homogeneous in this respect.

Our analyses aimed to test whether the effects of classroom composition can be conceived as resulting from rational behaviour by both teachers and students as proposed by De Vos (1986; 1989). It was investigated if teachers set higher goals for their students in high-ability classes

than in low-ability classes. Although our analyses demonstrated that teacher goals exert a considerable effect on student achievement and that the teacher goals are strongly dependent on a student's individual IQ, the evidence clearly failed to confirm the assertion that classroom composition effects result from rational behaviour through the goals set by the teachers. The research outcomes did not reveal a significant effect of the average classroom IQ on the goals set for the individual students (see table 2), nor was the classroom average of the teacher goals found to affect student achievement in the way it was expected (see table 3). The effect of the average classroom IQ appeared even to increase after controlling for the teacher goals. On the other hand, it should be noted that the fundamental assumptions that teachers set higher goals for the intelligent students and that the students' achievements increase when they are faced with higher goals were clearly confirmed *at the individual level*. The present study, however, showed that teachers set quite diverging goals for the students in their class, whereas in the argument by De Vos the emphasis is on the effects of standards for the entire classroom.

The fact that in the present study the classroom composition effects could not be accounted for by the teacher goals may to some extent be due to the way they were measured. The items on which the index measuring the teacher goals is based (see section 5) may have encouraged the teachers to emphasize the variation in approaches they employed for the students in their class. Perhaps, if more specific and objective information about their goals (e.g. through detailed questions about the subject matter they wanted their students to master) had been available, the within classroom differences with respect to the teacher goals might have appeared smaller. However, the fact that substantial differences between classrooms with respect to the teacher goals were found (see table 2) suggests that the index is able to distinguish classes where high goals are set from classes with low goals, apart from detecting differences within the classrooms.

It should also be noted that the index employed in the present study expressed the teachers' maximum aspiration levels rather than their minimum standards. When focusing on the minimum levels of achievement the teachers demand from their students classroom effects on individual achievement may still be (partly) accounted for by the teacher goals. The findings reported by Van der Velden (1993) support this idea. In his analyses the effects of two kinds of teacher goals on student achievement were investigated. Both variables related to the goals for the classroom as a whole and not to individual students. The first variable resembles the operationalization of teacher goals that was employed in the present study quite closely. This variable was not found to affect achievement, which is consistent with the findings in this study, as we only detected an effect at the individual level. The other variable, however, which relates to the minimum goals the teacher sets for the classroom did produce a significant effect. These outcomes suggest that at the classroom level the minimum standards set by the teachers are important, whereas our findings indicate that at the individual level the maximum aspiration levels of the teachers enhance achievement. To gain a more clear understanding of the causes and effects of teacher goals further (and probably quite elaborate) research is required.

More accurate measures of teacher goals may be obtained by asking teachers to indicate whether they want a certain student to master some well-defined subject matter. Such questions might even be asked about specific test-items. Separate questions might be formulated to measure the minimum standard and the maximum aspiration level a teacher sets for each student. Explicit questions could also be asked about the goals and standards at the classroom level. The present study was mainly focused on the maximum aspiration levels the teacher had set for each student with respect to not very specifically defined subject matters and it was assumed that the classroom mean of the individual goals would correctly reflect the goals the teachers had set for their class as a whole. Another issue that could not be addressed in the present study refers to the relation between the teacher goals and their actual behaviour in the classroom. We assume, however, that the information provided by the teachers presents a valid picture of their intentions. Furthermore, it should be pointed out that even when teacher goals and their behaviour are not clearly related, the goals may still have an important effect on achievement. Teacher goals reflect what they want their students to achieve and it seems plausible that teachers are able to attain a chosen goal via different approaches.

Our study has nevertheless demonstrated that a teacher may choose quite different goals for the students in his/her class. Even if the way the teacher goals were measured tends to exaggerate the differences within classes somewhat, which is for the moment no more than an unconfirmed suspicion, this conclusion seems unavoidable. This inevitably limits the extent to which teacher goals can account for classroom composition effects. Possibly, the interactions and competition among classmates may offer more valid explanations of classroom composition effects than the idea of differential instruction. It should be noted, though, that teachers may influence the interaction and competition among classmates, so that teacher behaviour may still account to some extent for the effects of classroom composition. In the present study the reinforcement of the inequalities between students which results from homogeneous grouping could certainly not be attributed to the goals set by the teachers. On the contrary, the effect of the classroom average IQ was found to increase after controlling for the teacher goals.

NOTES

1. The authors would like to thank Hennie Brandsma, Anja Knuver and Gerry Reezigt for providing part of the data (the "national sample", see section 5) that were analyzed in the present study.
2. Rath (1991, pp. 177-238) describes a case of two elementary schools in Rotterdam, housed in the same building, but one with a population of predominantly foreign children (90%) and the other with a student population of predominantly Dutch children (60%). The parents of foreign origin frequently protested against this segregation.
3. Most of these students originate from Surinam, Turkey, Morocco, Aruba or the Dutch Antilles (mainly Curaçao).
4. Although the test consists of 58 language items and only 30 mathematics items, the mathematics items are not underrepresented. The composite achievement score was obtained by computing the unweighted average of the (standardized) scores on the language and mathematics sub-tests.
5. The output measure was based on information about the school careers in secondary education of each school's students.
6. It should be noted that the classification of high, low and medium output was different for the three categories (high, low or medium input). Schools with an "average" output were included in the sample, but these were predominantly high-input schools with a relatively low output or low-input schools with a high output.
7. For the original Dutch formulations see Van der Hoeven-Van Doornum (1990, pp. 107-108).
8. This can be illustrated by considering the figures pertaining to the effect of "Standard deviation IQ" in the regional sample. Given the size of the standard error for this regression coefficient (.046) only an effect larger than .090 would be considered statistically significant, because only coefficients which are about twice as large as their standard error are significant for $1 < .05. If the actual size of the effect is smaller than .090, the chances to obtain a statistically significant result are less than 50 %. In other words: the power of the test is limited when it comes down to detecting small effects.
9. If standardized coefficients are preferred the product of the z-score transformations may serve as the interaction term (see section 6.2).

REFERENCES

Beckerman, T.M. & Good, T.L. (1981). The classroom ratio of high and low aptitude and its effect on achievement. *American Educational Research Journal, 18,* 317-327.

Blank, J.L.T., Boef-van der Meulen, S., Bronneman-Helmers, H.M., Herweijer, L.J., Kuhry, B. & Schreurs, R.A.H. (1990). *School en schaal* (School and scale). Rijswijk/Den Haag: Sociaal en Cultureel Planbureau/VUGA.

Blok, H. (1992). De grootte van het schooleffect in het basisonderwijs; een analyse op basis van vijf jaar Eindtoets Basisonderwijs. (The size of the school effect in elementary education; an analysis based on five years Final Elementary Education Test). *Tijdschrift voor Onderwijsresearch, 17,* 343-354.

Bosker, R. & Hofman, A. (1987). Dimensies van schoolkwaliteit. De algemene en milieuspecifieke invloed van scholen op prestaties en het keuzegedrag van leerlingen (Dimension of school quality. The general and background specific effect of schools. In J. Scheerens & W.G.R. Stoel, (eds.) *Effecti-*

viteit van onderwijsorganisaties (Effectiveness of educational organizations). Lisse: Swets & Zeitlinger.

Brandsma, H.P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs* (Characteristics of elementary schools and the quality of education; Dissertation, with a summary in English), Groningen: RION.

Brandsma, H.P. & Knuver, A.W.M. (1988). Organisatorische verschillen tussen basisscholen en hun effect op leerprestaties (Organizational differences between elementary schools and their effects on student achievement). *Tijdschrift voor Onderwijsresearch, 13*, 201-212.

Cronbach, L. (1987). Statistical tests for moderator variables: Flaws in analysis recently proposed. *Psychological Bulletin, 102*, 414-417.

Dar, Y. & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal, 23*, 357-430.

Duke, J.B. (1993). Estimation of the network effects model in a large dataset. *Sociological Methods and Research, 21*, 465-481.

Erbring, L. & Young, A.A. (1979). Individuals and social structure: Contextual effects as endogeneous feedback. *Sociological Methods and Research, 7*, 396-430.

Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education, 59*, 185-198.

Gamoran, A. (1992). Is ability grouping equitable? *Educational Leadership, 50* (2), 11-16.

Guldemond, H., Haanstra, F., Laarhoven, P. van & Vries, A.M. de (1987). *Effecten van heterogeen groepeeren in het voortgezet onderwijs, een voorstudie voor onderzoek in Nederland* (Effects of heterogeneous grouping in secondary education, a preliminary study for research in the Netherlands). Groningen: RION.

Grift, W. van de & Akkermans, W. (1991). Onderwijskundig leiderschap en leerprestaties in het basisonderwijs (Educational leadership and student achievement in elementary education) *Tijdschrift voor Onderwijsresearch, 16*, 244-257.

Hallinan, M.T. (1987). Ability grouping and student learning. In M.T. Hallinan (Ed.): *The social organization of schools* (pp. 41-69). New York/London: Plenum.

Van der Hoeven-van Doornum, A.A. (1990). *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen.* (Effects of teacher expectations and aspiration levels on school careers; Dissertation, with a summary in English) Nijmegen: ITS.

Jaccard, J., Turrisi, R. & Wan, C.K. (1990). *Interaction effects in multiple regression.* Newbury Park/ London/New Delhi: SAGE Publications.

Jungbluth, P. (1993). Pygmalion and the effectiveness of 'black' schools; teachers' stereotypes and hidden goal differentiations towards ethnic minorities. *Tijdschrift voor Onderwijsresearch, 18*, 99-110.

Kerckhoff, A.C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review, 51*, 842-858

Knuver, J.W.M. (1993). *De relatie tussen klas- en schoolkenmerken en het affectief functioneren van leerlingen* (The relation between class and school characteristics and the affective functioning of the students; Dissertation, with a summary in English) Groningen: RION.

Kulik, C.C. & Kulik, J.A. (1982). Effects of ability grouping on secondary school : a meta-analysis of evaluation findings. *American Educational Research Journal, 19*, 415-428.

Kulik, C.C. & Kulik, J.A. (1984). Effects of ability grouping on elementary school pupils: a meta-analysis. *Paper presented at the annual meeting of the American Psychological Association, Toronto.*

Leiter, J. (1983). Classroom composition and achievement gains. *Sociology of Education, 56*, 126-132.

Link, C.R. & Mulligan, J.G. (1991). Classmates' effects on black student achievement in public school classrooms. *Economics of Education Review, 10*, 297-310.

Maas, C.J.M. (1992). *Probleemleerlingen in het basisonderwijs.* (Pupils with learning and behavioral problems in elementary education; Dissertation, with a summary in English) Utrecht: ICS.

Meijnen, G.W. (1984). *Van zes tot twaalf. Een longitudinaal onderzoek naar de milieu- en schooleffecten van loopbanen in het lager onderwijs* (From six to twelve. A longitudinal study into the school effects and socio-economic background effects in elementary education). Harlingen: Flevodruk.

Mensen, T. & Guldemond, H. (1987). Groepseffecten op individuele leerprestaties (Classroom effects on individual achievement). *Mens en Maatschappij, 62*, 176-199.

Oakes, J. (1985). *Keeping track. How schools structure inequality.* New Haven: Yale University Press.

Prosser, R., Rashbash, J. & Goldstein, H. (1991). *ML3 Software for Three-level Analysis.* London: Institute of Education, University of London.

Rath, J.C. (1991). *Minorisering: de sociale contructie van "etnische minderheden"* (Minorization: the construction of "ethnic minorities"; Dissertation with a summary in English). Amsterdam: SUA

Reezigt, G.J. (1993). *Effecten van differentiatie op de basisschool.* (Effects of within classroom differentiation in elementary schools; Dissertation, with a summary in English). Groningen: RION

Resh, N. & Dar, Y. (1992). Learning segregation in junior high-schools in Israel: Causes and Consequences. *School Effectiveness and School Improvement, 3*, 272-292.

Rowan, B. & Miracle, A.W. (1983). Systems of ability grouping and the stratification of achievement in elementary schools. *Sociology of Education, 56*, 133-144.

Slavin, R.E. (1987). Ability grouping and achievement in elementary schools: a best-evidence synthesis. *Review of Educational Research, 57*, 293-336.

Slavin, R.E. (1990). Ability grouping and achievement in secondary schools: a best-evidence synthesis. *Review of Educational Research, 60*, 471-499.

Velden, L.F.J. van der (1993). The effects of teachers' attitudes and practices on the achievements of dutch low SES pupils. *Tijdschrift voor Onderwijsresearch, 18*, 85-98.

De Vos, H. (1986). Een model van het interactiesysteem van de schoolklas (A model of the interaction system of the classroom). *Mens en Maatschappij, 61*, 147-168.

De Vos, H. (1989). A rational choice explanation of composition effects in educational research. *Rationality and Society, 1*, 220-239.

De Vries, A.M. (1992). *Hoe breder hoe beter? De effecten van heterogeen groeperen in het voortgezet onderwijs in Nederland* (The broader the better? The effects of heterogeneous grouping in secondary education in the Netherlands; Dissertation with a summary in English). Groningen: RION.

Weide, M.G. (1993). Effective provisions for ethnic minorities in primary schools. *Tijdschrift voor Onderwijsresearch, 18*, 74-84.

Werf, M.P.C. van der & Weide, M.G. (1991). Effectief onderwijs voor allochtone leerlingen (Effective education for minority students). *Tijdschrift voor Onderwijsresearch, 16*, 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 Dutch Educational Priority Policy in the school and in the classroom. The first measurement in elementary education). Groningen/Nijmegen: RION/ITS.

Willms, J.D. (1986). Social class segregation and its relationship to pupils'examination results in Scotland. *American Sociological Review, 51*, 224-241.

Manuscript submitted: March 21, 1994
Accepted for publication: July 12, 1994

APPENDIX

Univariate statistics at the individual and classroom level

Variable	Mean	Median	Std. Dev.	Range	Minimum	Maximum
1. National Sample (3993 students, 221 classes)						
Individual Achievement Score	20.562	21.111	5.205	25.44	4.556	30.000
Class Mean Achievement	20.188	20.475	3.129	15.643	10.308	25.951
Individual SES	00.000	-0.371	6.695	25.000	-11.871	13.129
Class Mean SES	-0.304	-0.423	3.737	20.196	-11.167	9.029
Individual IQ	11.494	11.500	1.803	12.667	5.000	17.667
Class Mean IQ*	11.407	11.505	0.819	5.682	8.083	13.766
Standard Deviation IQ*	1.671	1.646	0.357	1.956	0.788	2.744
2. Regional Sample (698 students, 57 classes)						
Individual Achievement Score	534.447	536.000	10.801	49.000	501.000	550.000
Class Mean Achievement	533.240	534.143	7.197	31.069	514.556	545.625
Individual SES	00.000	-3.489	10.046	38.190	-11.599	26.591
Class Mean SES	-.545	-2.024	6.003	26.881	-8.764	18.117
Individual IQ	12.868	13.140	2.499	15.000	3.430	18.430
Class Mean IQ	12.754	12.751	1.357	5.964	9.126	15.090
Standard Deviation IQ	2.143	2.153	0.703	3.976	0.099	4.075
Individual Teacher Goals	3.607	3.780	0.715	4.000	1.000	5.000
Class Mean Teacher Goals	3.535	3.560	0.375	2.320	2.195	4.515

* These figures deviate from the ones in table 4. The figures in table 4. were obtained as follows: The classroom aggregates were disaggregated to the individual students, whereas in the appendix the figures relate to the frequency distributions across classes. Example: the average of the class mean IQ in the appendix is lower than in table 4, because the class mean IQ is lower in small classes. When computing the statistics in the appendix that relate to classroom aggregates across all classes, large or small, received the same weight. In this sense small classes are overrepresented.

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, pp. 63-78

Denominatie en effectiviteit van schoolbesturen

Roelande Hofman, Wijnand Hoeben en Henk Guldemond
RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen[1]

ABSTRACT

This article addresses the relative influence of schools' denomination (Catholic, Protestant, Public) and of schools' boards on school effectiveness. Existing data from a sample of 114 Dutch primary schools have been analyzed with multilevel variance component analysis to define a school's effects as its added value in learning gains after controlling for prior knowledge, IQ, SES, gender and ethnicity of its individual students. Two different regression analyses of the data on school level – the second analysis including denomination as a last step – demonstrate that its superficial influence – appearing in the first analysis using denomination as a covariate – may be explained by variables covering the schools' process, input and board. LISREL-analyses demonstrate that all influences of denomination on the effectiveness of schools are in fact mediated by variables covering the boards, which influences in turn are direct as well as indirect; their indirect influence is mediated by the process variable: evaluation of pupils' attainment.

1. INLEIDING

Begin 1993 is een proefschrift verschenen (Hofman, 1993a) waarin werd aangetoond dat kenmerken van het bestuur van een school bijdragen aan de schoolresultaten van de leerlingen op de desbetreffende school. Die additionele bijdrage bleef bestaan ook nadat is gecontroleerd voor eerdere prestaties, intelligentie, sociaal-economisch milieu en sekse van de leerlingen, nadat is gecontroleerd voor de invloed van contextvariabelen als denominatie, en nadat is gecontroleerd voor variabelen van schoolprocessen en van schoolinput. Met name de invloed die aan interne geledingen op school wordt gegeven op bestuursbeslissingen, heeft na de genoemde controles een significante invloed op de nettoleerwinsten voor taal en rekenen van de leerlingen. Schoolbesturen kunnen derhalve effectief zijn.

In perspublicaties en discussies is aandacht besteed aan de rol van de denominatie van de school in de effectiviteit van het bestuur. Er zijn vragen gesteld waarom scholen met een rooms-katholieke of een protestants-christelijke signatuur effectiever zijn als het gaat om leerwinsten bij taal, waarom scholen van rooms-katholieke signatuur effectiever zijn als het gaat om leerwinsten bij rekenen, en waarom scholen van rooms-katholieke of openbare signatuur effectiever zijn als het gaat om de schoolbeleving van de leerlingen. In het verlengde van de hypothese van Dronkers (1992) bestaat bij velen de neiging om antwoorden te geven in termen van bestuurskenmerken.

Dergelijke interpretaties passen niet bij de in het proefschrift gehanteerde analyse-strategie. In een stapsgewijze regressie zijn de effecten van denominatie immers in een eerste stap met denominatie als covariaat meegenomen, daarna de effecten van schoolprocesvariabelen en van schoolinput. De effecten van kenmerken van schoolbesturen zijn als laatste stap in de regressieanalyse ingevoerd. De effecten van de invloed van interne geledingen op bestuursbeslissingen gaan dan ook uit boven de effecten van de andere variabelen die in eerdere stappen zijn geconstateerd. De analyse-strategie om effecten van schoolbesturen vast te stellen is zodoende conservatief. Dit moet het doel om zeker te zijn dat de geconstateerde effecten inderdaad bestuurseffecten zijn in plaats van effecten van samen met bestuurskenmerken variërende andere kenmerken van scholen.

[1] Adres: Postbus 1286, 9701 BG Groningen.

Om na te gaan of invloeden van denominatie op leerwinsten misschien lopen via kenmerken van schoolbesturen, dan wel via kenmerken van schoolinput of schoolprocessen, zijn andere analysestrategieën nodig. In deze secundaire analyse van het datamateriaal, waarop het proefschrift is gebaseerd, wordt nagegaan of de effecten van denominatie kunnen worden verklaard via de invloed van dichter bij het primaire proces gelegen variabelen uit het onderzoeksmodel. Deze variabelen komen overeen met Dronkers (1992) verklaringen met behulp van het academisch klimaat en de gemeenschap van ouders, leraren en schoolbesturen. De probleemstelling die hier wordt onderzocht, kan als volgt worden geformuleerd:

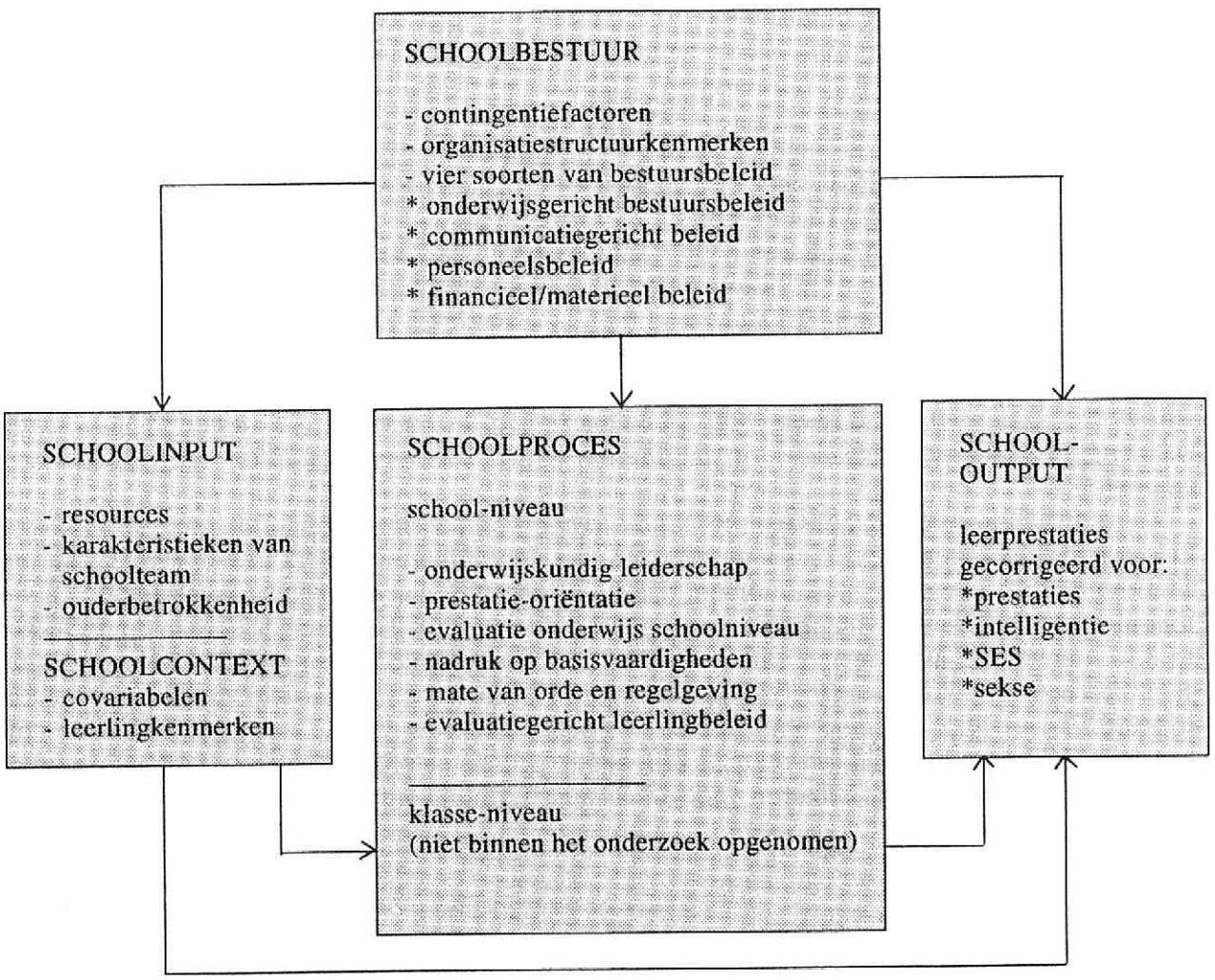
Is de gevonden samenhang tussen denominatie en effectiviteit van basisscholen te herleiden tot invloeden van variabelen die dichter bij primaire processen op school liggen, waaronder ook begrepen de bestuurskenmerken?

2. ONDERZOEKSMODEL, KERNBEGRIPPEN EN STEEKPROEF

Het onderzoeksmodel

In het onderzoek is een model voor de bijdrage van schoolbesturen aan de effectiviteit van basisscholen ontwikkeld dat een uitwerking is van het integrale schooleffectiviteitsmodel van Scheerens (1990). In Figuur 1 wordt dit model gepresenteerd.

Figuur 1: Onderzoeksmodel voor bijdrage van schoolbesturen aan schooleffectiviteit.



De kernbegrippen

De kernbegrippen die daarbij zijn gehanteerd zijn betreffen (a) de zogenaamde contingentiefactoren. Hierbij gaat het om indicatoren voor drie contingentiefactoren uit de lokale omgeving van schoolbesturen voor openbaar en voor bijzonder onderwijs namelijk bestuurlijke schaal, omgevingsinvloed en externe ondersteuning, die de interne structuur en het functioneren van schoolbesturen beïnvloeden. In bijlage I worden de psychometrische statistieken van de gehanteerde schalen gepresenteerd. De tweede groep factoren betreft (b) de organisatiekenmerken van schoolbesturen. Ook hier gaat het om indicatoren die drie centrale begrippen representeren namelijk: interne bestuursdifferentiatie, deskundigheid en cohesie en tenslotte planmatigheid van het bestuur. In tabel 2 van de bijlage worden de psychometrische kenmerken van deze indicatoren gepresenteerd. Het (c) bestuursbeleid is gebaseerd op een factoranalyse over een (uitputtende) indeling van de mogelijke bestuursactiviteiten. Op basis van beschikbare empirische gegevens is gekozen voor een indeling naar zes taakgebieden die de realiteit van het handelen van schoolbesturen het meest recht doet. Die zes taakgebieden representeren tezamen het bestuursbeleid. Per taakgebied zijn twee schalen geconstrueerd: een frequentieschaal en een taakopvattingschaal (een meer fiatterende of meer beleidsvormende bestuursopvatting per deeltaak). De resultaten van die schaalconstructie, de psychometrische kenmerken en validiteitsgegevens worden in bijlage I gepresenteerd. De mate van interne consistentie van de schalen is voldoende tot hoog. De schoolbesturen zijn het meest actief op het terrein van de interne en externe contacten en op beheersmatige terreinen zoals het financieel-materieel beleid en het personeelsbeleid. De besturen zijn het minst actief op de onderwijsgerichte taakgebieden zoals het onderwijs- en organisatiebeleid. Dit komt overeen met uitkomsten uit ander Nederlands onderzoek naar het functioneren van schoolbesturen (Detering & Kalkman, 1985; Braster & Leune, 1986; Buurke, 1988; Van Esch 1988) èn bevestigt de uitkomsten van een case-study die voorafgaande aan dit onderzoek is uitgevoerd.

Deze deeltaken zijn aan een factoranalyse onderworpen hetgeen resulteerde in vier bestuursdomeinen. In totaal wordt via deze factorstructuur 75% van de variantie in bestuursbeleid gebonden. Deze vier bestuursdomeinen representeren in het vervolg tezamen het bestuursbeleid.

Bij het cluster schoolinput is de centrale variabele de mate waarin scholen beschikken over extra formatie. Dit betekent in feite een meting van de leerlinginstroom van de school omdat scholen met relatief veel leerlingen met een lage sociaal-economische herkomst of van allochtone afkomst binnen het onderwijsvoorrangsbeleid extra formatie krijgen toebedeeld.

Tenslotte het cluster schoolprocesvariabelen. Het betreft hier een zevental indicatoren van schooleffectiviteit. In bijlage I worden ook daarvan de psychometrische statistieken gepresenteerd.

De steekproef

Het onderzoek betreft een databestand waarin scholen en leerlingen zijn onderzocht. Van dat bestand zijn de besturen benaderd en in totaal 146 besturen gaven hun medewerking. De respons bedroeg 65%. De data hebben aldus betrekking op ongeveer 2650 leerlingen, 146 scholen en schoolbesturen. Belangrijk is op te merken dat de informatie van de schoolbesturen telkens is bevraagd voor één, bij naam genoemde, basisschool van dat bestuur. Het analyseniveau dat is gehanteerd betreft de school. In de volgende paragraaf wordt daarop nader ingegaan.

3. ANALYSESTRATEGIEËN

Voor het nagaan of invloeden van denominatie op leerwinsten voor rekenen en taal lopen via kenmerken van schoolbesturen, van schoolinput of van schoolprocessen, zullen twee analysestrategieën worden gebruikt: stapsgewijze regressie-analyse en LISREL-analyse.

Bij de toepassing van deze analyse-strategieën moet rekening worden gehouden met de hiërarchische structuur van de school als organisatie: leerlingen behoren tot een school en deze

school valt onder een schoolbestuur. De eenheid van organisationele analyse is in deze studie de school.

In een stap voorafgaand aan de twee analysestrategieën wordt de effectiviteit van elk van de scholen bepaald op de gekozen outputmaten namelijk rekenen en taal. Die schooleffectiviteit is gemeten als de schoolspecifieke afwijking van het steekproefgemiddelde voor taalprestaties, rekenprestaties en schoolbeleving in groep 8. Dit nadat is gecontroleerd voor de beginprestaties in groep 7 en de schoolspecifieke verschillen in de instroom (van de afzonderlijke leerlingen) met betrekking tot IQ, milieu, etniciteit en sekse. Een bovengemiddelde score op deze effectiviteitsmaat indiceert dat de scholen er beter in slagen hun leerlingen hogere prestaties te laten behalen dan verwacht mag worden op basis van de individuele achtergrondkenmerken van deze leerlingen. Deze maat is gebaseerd op geaggregeerde, gestandaardiseerde, individuele leerlingscores. Deze mate van schooleffectiviteit is berekend met behulp van het multi-level analyseprogramma VARCL (Longford, 1986). Het resultaat van deze variantie-componenten-analyse wordt uitgedrukt in 'posterior means' per school en per outputmaat. De outputmaten geven de prestaties weer van leerlingen op verschillende door het CITO geconstrueerde toetsen op het terrein van rekenvaardigheid en taalvaardigheid gecorrigeerd voor de genoemde individuele leerlingkenmerken en maken zodoende een 'faire' vergelijking van scholen mogelijk, namelijk een vergelijking van hun toegevoegde waarde.

Omdat in het gebruikte databestand twee prestatiemetingen zijn uitgevoerd is het mogelijk het 'zuivere' effect van scholing, de leerwinst, vast te stellen (Hofman & Hofman, 1988). Dit betekent dat bij de uiteindelijke geconstrueerde effectiviteitsmaat op schoolniveau ook rekening wordt gehouden met de eerdere leerprestaties van één schooljaar terug. Bosker (1991) spreekt in dit verband van de netto leerwinst die leerlingen in één schooljaar boeken. De geconstrueerde effectiviteitsindices per school – de netto leerwinst voor rekenen en taal als basis voor de berekening van de toegevoegde waarde van een school – vormen de criterium-variabelen in het onderzoeksmodel. De resultaten van deze eerste stap worden hier eerst besproken alvorens wordt ingegaan op de twee analyse-strategiën.

Verdeling van variantie-componenten

Met behulp van VARCL analyses is het mogelijk variantiecomponenten op verschillende niveaus (in ons geval leerling en school/bestuur) uiteen te leggen en te presenteren als proporties van de totale variantie (Aitkin & Longford, 1986, Van Eeden & Meijnen, 1990). In tabel 1 wordt een overzicht gepresenteerd van de verdeling van de variantie-componenten in het lege model, in het leerwinstmodel (waarbij de prestaties van het schooljaar 1987 als covariabele worden gehanteerd op de prestaties in 1988) en in het uiteindelijk voor de schooleffectiviteitsindices gebruikte leerwinst en covariatenmodel. Dit laatste model betreft een verdeling van variantiecomponenten over leerlingniveau en schoolniveau met eerdere schoolprestaties en leerlingken-

Tabel 1. Proportieverdeling variantie-componenten voor taal en rekenen.

	leeg model	model leerwinst	leerwinst en covariaten model
Rekenen			
variantie-component:			
leerlingniveau	.728	.450	.366
schoolniveau	.283	.138	.116
Taal			
variantie-component			
leerlingniveau	.792	.396	.345
schoolniveau	.219	.097	.083

merken als covariaten. De proportie variantie per model wordt voor de criteriumvariabelen rekenen en taal in Tabel 1 gepresenteerd.

Uit deze multi-level-analyses van de leerwinstscores voor taal en rekenen vallen een aantal zaken op. Allereerst blijkt voor het leergebied taal dat het netto percentage tussenschoolse variantie dat resteert na controle voor de taalprestaties van 1987 en voor de covariabelen, 8% bedraagt. De tussenschoolse variantie is groter voor het leergebied rekenen, namelijk 12%. Een relatief groot deel van de variantie wordt gebonden door de rekenscore en taalscore in het voorgaande jaar. Het verschil in tussenschoolse variantie tussen rekenprestaties en taalprestaties is in overeenstemming met resultaten van ander onderzoek. In de literatuur wordt als verklaring voor dit verschil veelal gewezen op het schoolgebonden karakter van het leren rekenen en het sterk milieugebonden karakter van het aanleren van taal (Mortimore, e.a. 1988; Bosker & Hofman, 1987; Madaus, 1980; Meijnen, 1984). In Nederland en in het buitenland vastgestelde percentages tussenschoolse variantie voor rekenen en taal stemmen overeen met de hier gepresenteerde cijfers (vgl. ook Creemers & Scheerens, 1991).

Analysestrategieën ter beantwoording van de probleemstelling

Na deze constructie van de criteriumvariabelen op schoolniveau (rekenen en taal) is de eerste analysestrategie een stapgewijze strategie voor regressie-analyse, zoals ook in het proefschrift gehanteerd, maar dan met een andere volgorde van inclusie van de clusters (context, bestuur, input, proces) uit het onderzoeksmodel. De eerste stap is het invoeren van variabelen van het schoolproces (let wel: klasse- en instructiekenmerken zijn niet opgenomen in het databestand). De tweede stap is het invoeren van variabelen van de schoolinput. De derde stap is het invoeren van variabelen van het schoolbestuur. De laatste stap is het invoeren van overige variabelen van de schoolcontext, zoals denominatie, schoolgrootte, SES-compositie, urbanisatiegraad e.d.

Er zijn bij een analyse van de rol van denominatie enkele argumenten voor deze, in plaats van de in het proefschrift gekozen volgorde van stapsgewijze 'inside-out' regressie-analyse. Op deze manier wordt echt van binnen naar buiten gewerkt, in plaats van uit overwegingen van zorgvuldige toetsing eerst covariaten van de context in de regressie-analyse in te voeren. De in dit artikel gehanteerde analysestrategie werkt zodoende naar analogie van een ui. Een model van causale ketens, waarbij de buitenste schil van oorzaken van invloed is op factoren in de volgende schil, die vervolgens weer van invloed zijn op factoren in een verder naar binnen gelegen schil, enzovoort. Rechtstreekse invloed van variabelen in de buitenste schil op variabelen in de kern is een invloed die aanvullend is op of in interactie is met variabelen in meer naar binnen gelegen schillen.

Een volgend argument voor de hier gehanteerde strategie is dat op deze manier waarschijnlijk een grotere bijdrage van variabelen in meer naar binnen gelegen schillen zal worden gevonden. Voor een conservatieve vaststelling van het effect van schoolbesturen (zoals in het proefschrift beschreven) is zoiets niet handig. Voor een verklaring en interpretatie van de invloed van denominatie is het echter juist wel nodig. Een dergelijke analyse-strategie kan de effecten van denominatie mogelijk verklaren met behulp van variabelen die dichter bij het primaire leerproces liggen. Misschien nog belangrijker dan een verklaring op zichzelf is dat dergelijke variabelen gemakkelijker beïnvloedbaar zijn dan de denominatie van een school.

De tweede analysestrategie bestaat uit het expliciteren van causale paden die van buiten naar binnen lopen, dat wil zeggen van contextuele en bestuurskenmerken naar primaire processen op school en naar leerwinsten. Als de invloed van contextuele en bestuurskenmerken op leerwinsten loopt via kenmerken van schoolprocessen, dan wordt dit in LISREL-analyses zichtbaar. Als de invloed van context- en bestuurskenmerken niet via schoolprocessen verloopt, dan worden de statistische verbanden tussen context- en bestuurskenmerken enerzijds en effectiviteit anderzijds causaal gezien niet weg verklaard met behulp van deze analyse-strategie.

LISREL-analyses zijn bij uitstek geschikt om een 'inside-out' analyse uit te voeren in een causale structuur waarbij naar analogie van een ui factoren in een buitenste schil van invloed zijn op factoren in meer naar binnen gelegen schillen, die op hun beurt weer factoren in de kern beïnvloeden. Met deze analyses kunnen paden waarlangs oorzakelijkheid verloopt, worden

getoetst. De redenen waarom denominatie zoveel invloed heeft op schooleffectiviteit, kunnen zodoende binnen het databestand worden geanalyseerd. Nagegaan kan worden of deze redenen terecht worden gelegd bij verschillen in bestuur tussen openbare en confessionele scholen.

4. RESULTATEN VAN DE REGRESSIE-ANALYSES

De regressie-analyses worden in twee varianten gepresenteerd. Variant I is de analyse zoals eerder uitgevoerd (zie Hofman, 1993a). Variant II is de analyse met de hiervoor besproken andere volgorde van inclusie van variabelenclusters. Het verschil tussen variant I en variant II is gelegen in het moment van inclusie van de factor denominatie. Denominatie betreft twee variabelen namelijk openbaar versus katholiek en openbaar versus protestants-christelijk. In de eerste variant is de factor denominatie als eerste in de regressie-analyse opgenomen, terwijl in de tweede variant denominatie juist als laatste factor in de analyse wordt opgenomen. Zoals in paragraaf 1 al is aangegeven, wordt in variant I een analyse-strategie gehanteerd die erop gericht is netto effecten van schoolbesturen vast te stellen door uitzuivering van andere variabelen, waaronder begrepen de covariaten op schoolniveau. Variant II is er specifiek op gericht de effecten van denominatie op leerprestaties te verklaren met behulp van variabelen die dichter bij het primaire proces liggen.

In Tabel 2 worden de resultaten van de vergelijking van variant I en variant II gepresenteerd voor netto leerwinst rekenen. De significante predictoren worden per analyse-variant naast elkaar gepresenteerd met de hoeveelheid variantie die zij verklaren in de 2e criteriumvariabele. Bij de vergelijking van variant I met variant II valt allereerst op dat de verklaarde variantie van 42% in variant I terugloopt naar 40% in variant II. Verder doen zich in de predictoren van netto leerwinst rekenen bij een verandering in de inclusie-volgorde voor denominatie geen wijzigingen voor.

In samenhang met de verwachtingen vallen enige wijzigingen op in de verklaringskracht van de predictoren. De verklaringskracht van de schoolproces-variabelen voor de netto leerwinst bij rekenen is in de tweede variant totaal 24%, terwijl dit in de verklaarde variantie door die variabelen in de eerste variant 16% bedroeg. De door de factor 'invloed van schoolgeledingen op bestuursbeslissingen' verklaarde variantie bedroeg in de eerste variant 4% en loopt in de tweede variant op naar 6%. Het lijkt erop dat de verklaarde variantie van de oorspronkelijke factor denominatie volledig wordt geïntermedieerd door de schoolproces- en de bestuursvariabelen. Dit betekent dat het positieve verband tussen katholieke scholen en rekenprestaties verloopt via enerzijds de schoolprocesvariabele 'evaluatie van rekenprestaties' en in mindere mate

Tabel 2. Regressie-analyse variant I en variant II voor rekenen.

Netto leerwinst rekenen	VARIANT I Mult.R	Rsq	beta	VARIANT II Mult.R	Rsq	beta
1. Covariaten						
- denominatie: rooms-katholiek	.36	.13	.36			
2. Schoolprocesvariabelen						
- evaluatie leerlingprestaties	.47	.23	.33	.40	.16	.40
- nadruk op basisvaardigheden	.54	.29	.25	.49	.24	.29
3. Schoolinput						
- extra formatie	.61	.38	-.30	.58	.34	-.31
4. Schoolbestuur						
(a) bestuursbeleid						
(b) contingentiefactoren						
- invloed interne geledingen op bestuursbeslissingen	.65	**.42**	.22	.63	**.40**	.25

Tabel 3. Regressie-analyse variant I en variant II voor taal.

Netto leerwinst taal	VARIANT I Mult.R	Rsq	beta	VARIANT II Mult.R	Rsq	beta
1. Covariaten						
- denominatie: openbaar	.38	.14	-.13			
2. Schoolprocesvariabelen						
3. Schoolinput						
- extra formatie	.27	.07	-.27			
4. Schoolbestuur						
(a) bestuursbeleid						
(b) contingentiefactoren						
- invloed int.geledingen op bestuursbeslissingen	.45	**.20**	.18	.37	.14	.27
- bestuursvergaderingen met reguliere geledingen	.43	**.18**	.22			

via de 'nadruk op basisvaardigheden' en anderzijds via de bestuursvariabele 'invloed van interne geledingen op bestuursbeslissingen'.

Dezelfde analyse-strategie is toegepast voor netto leerwinst voor taal. In tabel 3 worden de resultaten gepresenteerd. Bij vergelijking van de twee varianten is wederom sprake van een terugloop in verklaarde variantie van 2%. Opmerkelijk is hier echter dat ook een mutatie in predictoren optreedt. In de tweede variant treedt bij het cluster school-input nu (net als bij rekenen) de extra formatie als voorspeller van leerwinst in taal op. Verrassend is dat er additioneel een contingentiefactor als predictor van leerwinst in taal naar voren komt. Bovendien is het opmerkelijk dat deze factor (namelijk de mate waarin bestuursvergaderingen worden gehouden tezamen met reguliere geledingen van de bestuurde school) ook inhoudelijk goed past bij de contingentiefactor 'invloed van interne geledingen op bestuursbeslissingen'. De invloed van die interne groeperingen op bestuursbeslissingen verloopt dan gedeeltelijk via die bestuursvergaderingen.

De conclusie die kon worden getrokken voor netto leerwinst rekenen, gaat in feite in nog sterkere mate op voor taal. Het negatieve verband tussen openbare scholen en taalprestaties wordt bemiddeld door enerzijds de schoolinput van de openbare school en anderzijds door bestuursvariabelen. De verklaringskracht van de bestuurvariabelen neemt toe van 6% tot 11%. Opmerkelijk is dat die bestuursvariabelen eenduidig op het vlak van de communicatie tussen bestuur en school liggen. Op grond van de gehanteerde analyse-strategie lijkt het aannemelijk dat de bestuursvorm van openbare scholen in enige mate een verklaringsgrond biedt voor het minder goed presteren van openbare scholen voor taal.

Nu aannemelijk is gemaakt dat de invloeden van denominatie op effectiviteit kennelijk via andere variabelen verlopen, wordt het interessant om te zien of deze resultaten worden bevestigd door de LISREL-analyses, en of die analyses weten aan te wijzen via welke variabelen de invloeden van denominatie verlopen.

5. CAUSALE MODELLEN: LISREL-ANALYSES

Voor LISREL-analyses van de relatieve invloeden van denominatie en van bestuurskenmerken zijn causale modellen nodig die de invloed van schoolbesturen van buiten naar binnen gaande expliciteren in de vorm van causale beïnvloedingspaden. Voor de onderhavige analyses houden we rekening met de beperkingen van het datamateriaal. Zoals voor integrale schooleffectiviteitsmodellen wordt bepleit (Scheerens, 1990), kunnen in de data blokken variabelen worden

onderscheiden. Allereerst met betrekking tot de context van een school, vervolgens met betrekking tot het bestuur van een school, met betrekking tot de input van een school, met betrekking tot schoolprocessen, en met betrekking tot de schooloutput. Ten overvloede, schooloutput wil hier dus zeggen de voor individuele leerlingkenmerken en leerprestaties uitgezuiverde schoolresultaten (zie paragraaf 3). Denominatie is weliswaar onderdeel van de contextvariabelen, maar wordt voor de doeleinden van de probleemstelling van deze secundaire analyse als een eigenstandig blok beschouwd.

De fundamentele veronderstelling van een van binnen naar buiten gaande analyse met behulp van integrale schooleffectiviteitsmodellen is, dat denominatie, context, bestuur en input geen rechtstreekse invloeden hebben op de schoolresultaten, maar invloeden die worden bemiddeld door schoolprocessen. Uit deze veronderstelling kunnen twee -meer of minder- complexe causale modellen worden gedistilleerd.

Het eerste model volgt bij de 'inside-out' analyse, die behoort bij integrale schooleffectiviteitsmodellen, strikt de analogie van de ui. Elke schil van kenmerken beïnvloedt slechts die kenmerken direct, die in de eerstvolgende schil worden onderscheiden. Invloeden op kenmerken in daarop volgende schillen zijn indirect; daarbijbehorende statistische verbanden zijn causaal gezien spurious (Figuur 2).

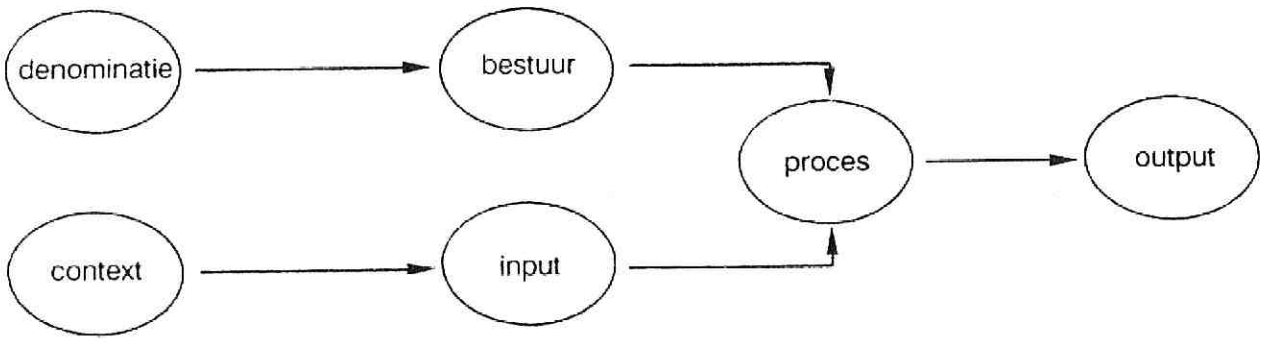


Fig. 2. Causaal model van integrale schooleffectiviteit met een minimum aan directe beïnvloedingslijnen.

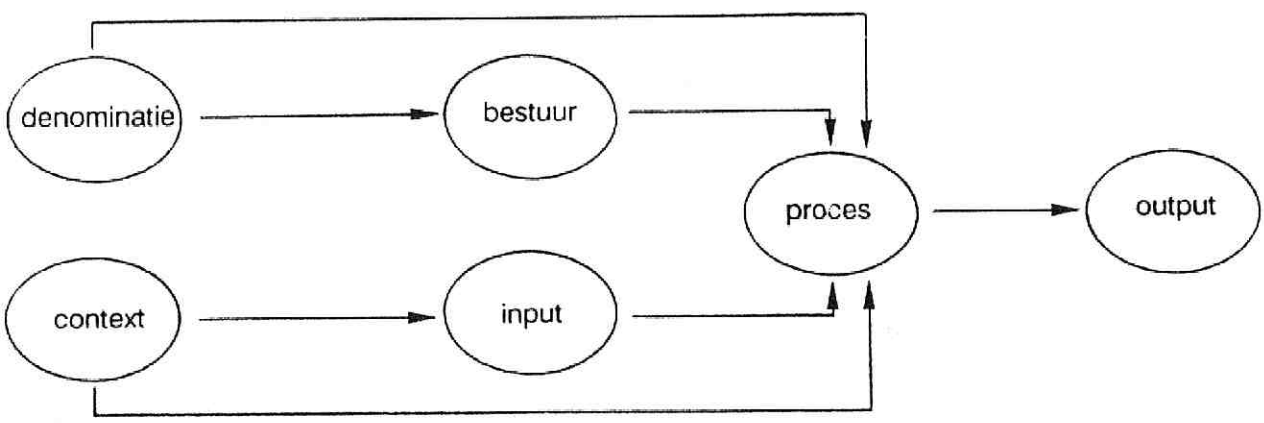


Fig. 3. Causaal model van integrale schooleffectiviteit met meerdere directe beïnvloedingslijnen op de procesvariabelen.

Het tweede model volgt minder strikt de gelaagdheid van de causale structuur van een 'inside-out' analyse. In het tweede model worden de beïnvloedingslijnen uit het eerste model verondersteld. Bovendien wordt verondersteld dat denominatie en context niet alleen een indirecte, maar ook een directe invloed hebben op de procesvariabelen (zie Figuur 3).

Rekening houdend met een ondergespecificeerdheid van deze causale modellen van integrale schooleffectiviteit kunnen twee varianten van een causale structuur worden gespecificeerd. Die staan telkens verder af van de veronderstellingen van een 'inside-out' analyse van integrale schooleffectiviteit. De eerste is een variant op het eerste model en veronderstelt dat in de data onvoldoende procesvariabelen zijn gemeten, c.q. dat de procesvariabelen onvoldoende hebben gemeten waarom het bij schooleffectiviteit gaat. Een gevolgtrekking is dat invloeden van input en van bestuur op de output niet alleen via de gemeten procesvariabelen verlopen, maar ook via ongemeten variabelen. In een secundaire analyse worden dan rechtstreekse invloeden van input en van bestuur op de output zichtbaar (Figuur 4).

De tweede variant, die nog verder af staat van de veronderstellingen van een 'inside-out' analyse van integrale schooleffectiviteit, veronderstelt dat in de data ook onvoldoende bestuursvariabelen zijn gemeten. Een gevolgtrekking is dat invloeden van denominatie en context op de

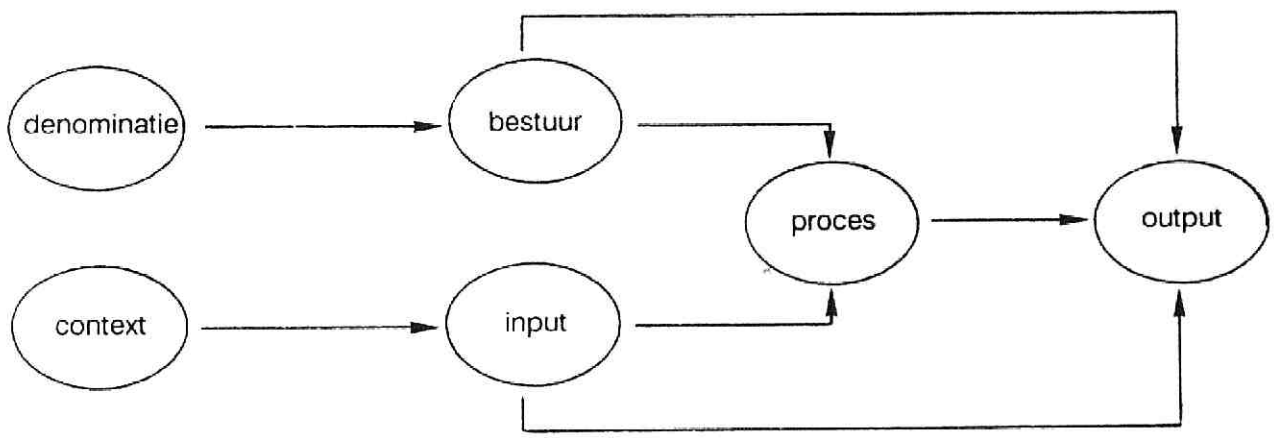


Fig. 4. Causaal model van schooleffectiviteit met directe beïnvloedingslijnen van bestuur en input op de output.

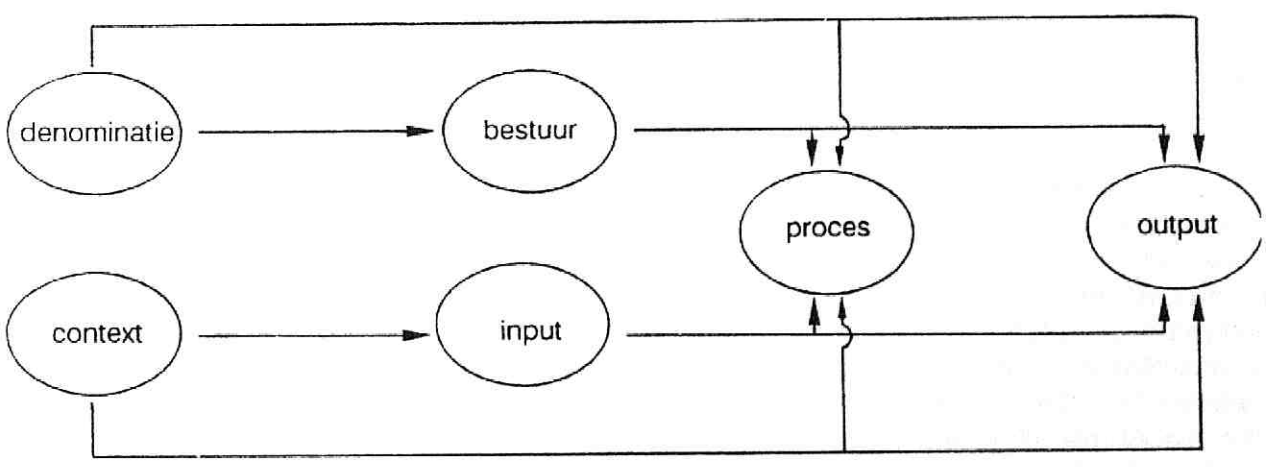


Fig. 5. Causaal model van schooleffectiviteit met directe beïnvloedingslijnen van alle variabelen op de output.

output niet alleen verlopen via de gemeten bestuurs-, input- en procesvariabelen, maar ook via ongemeten variabelen. In een secundaire analyse worden dan rechtstreekse invloeden van denominatie en context op de output zichtbaar. Deze variant is een combinatie van het tweede en het derde model (Figuur 5).

Bij de speurtocht naar passende LISREL-modellen zijn analyses uitgevoerd te beginnen met het eerste causale model (Figuur 1). De output, het schooleffect, is berekend als een latente variabele van netto leerwinst taal en netto leerwinst rekenen. Twee gegevenheden compliceerden de aanvankelijke analyses. Ten eerste was er het feit dat er een groot aantal indicatoren (in totaal 37) in het meetmodel was opgenomen, geclusterd rond de vijf blokken in de causale modellen (figuren 2 t/m 5). Ten tweede was het structurele model opgebouwd uit twee, grotendeels, van elkaar onafhankelijke 'sporen', respectievelijk het spoor van denominatie en bestuur, en het spoor van context en input. Deze gegevenheden hadden tot gevolg dat de geschatte variantie-covariantie matrix (Σ) een grote mate van discrepantie vertoonde met de geobserveerde matrix (S). Met andere woorden, de incongruentie tussen model en data vloeide voort uit de veronderstelling dat er geen samenhang bestaat tussen de clusters van context- en bestuursvariabelen enerzijds en denominatie en input-variabelen anderzijds. De 'modification-indices' gaven geen enkele aanleiding tot de veronderstelling dat er binnen dit model met een Chi-kwadraat van 429.47 bij 194 vrijheidsgraden ook maar enige verbetering van substantiële omvang te verwachten was. Ook modeluitbreidingen in de richting van de overige causale modellen zou geen betere passing opleveren.

In eerste aanleg is onderzocht in hoeverre een reductie per cluster van het aantal variabelen in het meetmodel uitkomst zou bieden. Per cluster werden de onderling laagste correlerende indicatoren verwijderd. Deze procedure leverde echter niet een dermate scherpe daling in de Chi-kwadraatwaarde op dat dit een duidelijke compensatie voor het verlies aan vrijheidsgraden inhield. Net zoals dat bij het meer uitgebreide meetmodel het geval was, konden ook hier geen substantiële verbeteringen in de passing van het model worden verwacht. Wel werd uit deze analyse duidelijk dat het cluster van context-variabelen kon worden aangemerkt als de voornaamste veroorzaker van de slechte passing met het model. Gezien de resultaten van de hiervoor gepresenteerde regressie-analyses wekt dat geen verbazing: ook daar spelen variabelen van de schoolcontext geen rol. Dit resultaat is niet verrassend omdat de geconstrueerde criteriumvariabele een uitzuivering van context-effecten inhoudt. Een volgende aanpassing van het gehanteerde model betreft aldus een volledige verwijdering van alle variabelen in het cluster van de schoolcontext.

Uit de uitgevoerde analyses blijkt dat het na deze aanpassing wel mogelijk is om een bij het causale model van figuur 2 passend LISREL-model te vinden (df=21, Chi^2=28.57, p=.125 en de GFI=.96; zie bijlage II). Inspectie van de modificatie-indices maakt duidelijk dat een verbetering van dit model mogelijk is door een direct pad te trekken van het cluster bestuur naar de outputmaten (een mengvorm van figuur 2 en figuur 4). Dit leidt tot een lichte verbetering van het model: de verklaarde variantie in de outputmaat loopt op van 19% tot 25%; de indices van de passing van het model zien er als volgt uit df=20, Chi^2=23.54, p=.263 en de GFI=.96). Dit is tevens het uiteindelijke model voor de verbanden tussen denominatie, schoolbestuur, procesvariabelen en schooleffectiviteit (Figuur 6).

De overall effecten op de outputmaat (leerwinst taal en rekenen) liggen voor denominatie het contrast van bijzonder versus openbaar in de orde van grootte van: -.88*.30*.34 + -.88*.20= -.265 en voor het bestuur in de orde van grootte van .30*.34 + .20=.302. De input van de school verklaart op schoolniveau nog .19*.19=.036. De effectgrootte van de procesvariabelen op de output is het grootst, namelijk .34 + .19=.53. Dit LISREL-model maakt duidelijk dat het effect van denominatie op de effectiviteit van basisscholen volledig loopt via het directe en het indirecte effect van het bestuur.

Bij denominatie moet worden opgemerkt dat sprake is van een negatief effect. In deze context betekent dit, dat sprake is van een negatieve invloed van het openbaar bestuur vergeleken met confessionele besturen op de effectiviteit van basisscholen. De bestuurskenmerken die er toe doen zijn dan naar gewicht (a) de mate waarin sprake is van bestuurvergaderingen met

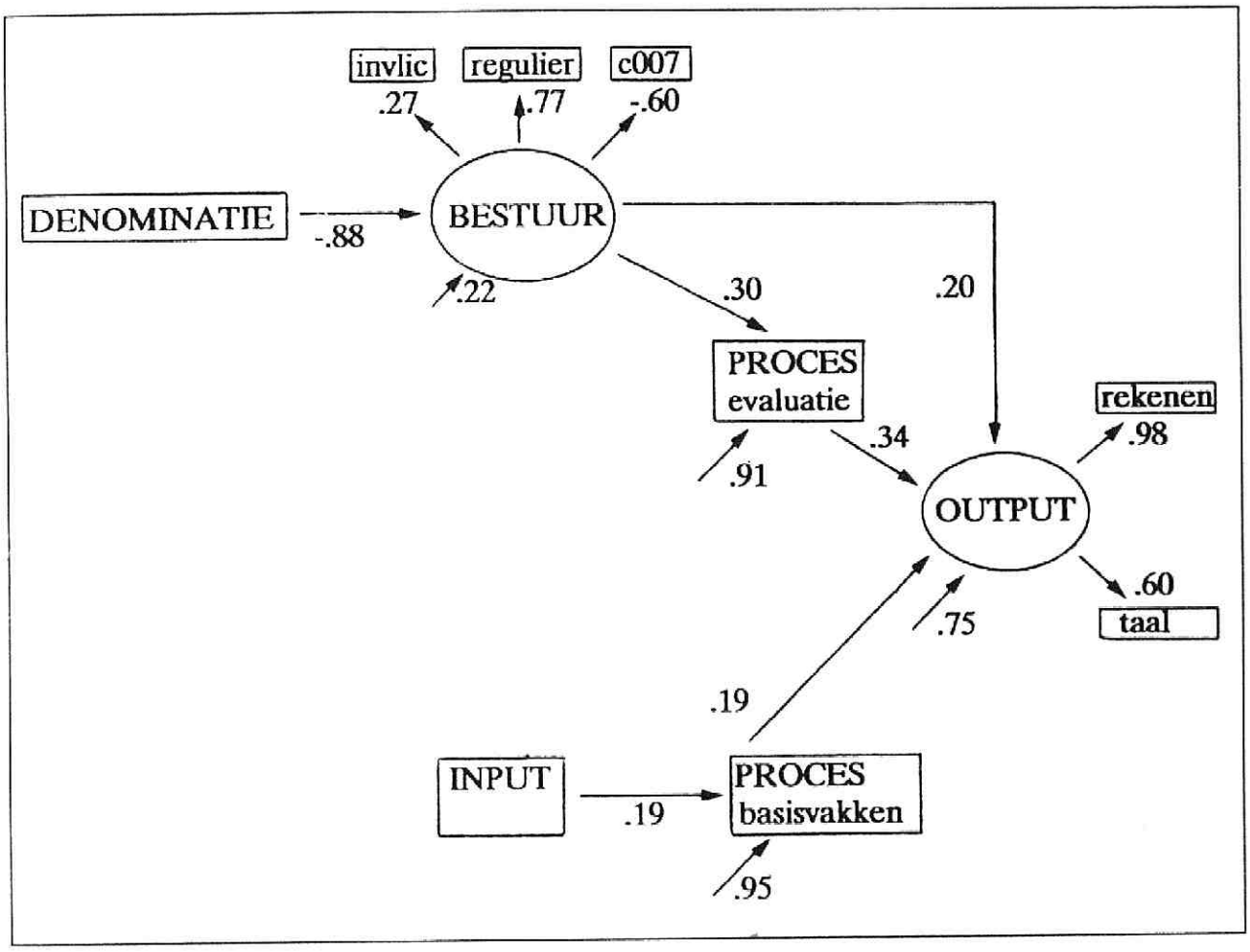


Fig. 6. Het uiteindelijk LISREL-model voor de samenhang van denominatie, bestuur, schoolproces en schooleffectiviteit.
(Chi2=23.54, df=20, p=.263 en GFI=.96).

reguliere geledingen in de school (vergadering; positief), (b) het aantal scholen dat onder het schoolbestuur ressorteert (schaal; negatief) en (c) tenslotte de mate waarin sprake is van invloed van interne geledingen op bestuursbeslissingen (invloed; positief).

6. CONCLUSIE EN DISCUSSIE

In deze paragraaf worden de resultaten nog eens belicht vanuit een methodische en een inhoudelijke invalshoek en vanuit de betekenis voor de praktijk van het besturen van scholen.

In een vergelijking van de regressie-analyses volgens de twee varianten wordt duidelijk dat het effect van denominatie in het databestand in belangrijke mate verloopt via de kenmerken van het schoolproces, van de schoolinput en van het bestuur. De gevonden samenhang tussen denominatie en effectiviteit van basisscholen kan dus grotendeels worden herleid tot dichter bij de primaire processen gelegen (manipuleerbare) variabelen. De regressie-analyses volgens de tweede variant verklaren echter wat minder variantie. Wanneer we de resultaten van de regressie-analyses vergelijken met die van de LISREL-analyses valt bovendien op dat bij de regressie-analyses gemiddeld meer variantie wordt verklaard dan bij de LISREL-analyses: lopend van 20% (18%) voor taal en 42% (40%) voor rekenen in de regressie-analyses naar 25% in het uiteindelijke LISREL-model. Het samen nemen van taal en rekenen in één outputmaat biedt minder verklaringskracht.

De uitkomsten van de LISREL-analyses specificeren de conclusie over de indirecte effecten van denominatie. Denominatie heeft in het uiteindelijke model alleen een direct effect op de bestuurskenmerken. Slechts via deze bestuurskenmerken (en de laatste invloed gedeeltelijk bemiddeld via evaluatie van leerlingprestaties) heeft denominatie invloed op de effectiviteit van basisscholen. De bestuurskenmerken die er toe doen zijn dan naar gewicht: (a) positief is de mate waarin sprake is van bestuursvergaderingen met reguliere geledingen in de school, (b) negatief is het aantal scholen dat onder het schoolbestuur ressorteert, en (c) positief is tenslotte de mate waarin sprake is van invloed van interne geledingen op bestuursbeslissingen. Het indirecte effect van denominatie op de effectiviteit van basisscholen is in overeenstemming met de oorspronkelijke verwachtingen (zie figuur 1). Het indirecte effect bevestigt de verklaring dat confessionele scholen gemiddeld effectiever zijn ten gevolge van de gemeenschap van ouders, leraren en schoolbesturen (Dronkers, 1992).

De overige effecten op de effectiviteit van scholen zijn bescheiden en afkomstig van een andere procesvariabele – de nadruk op basisvaardigheden – en via deze van de schoolinput. Dit gedeelte van de oorspronkelijke verwachtingen van figuur 1 is derhalve eveneens bevestigd.

De overige verwachtingen van figuur 2 zijn weerlegd. Zowel uit de regressie-analyses als uit de LISREL-analyses komt naar voren dat kenmerken van de schoolcontext, c.q. andere contextkenmerken dan bestuursvariabelen, geen invloed – noch direct, noch indirect – hebben op de effectiviteit van basisscholen. Al eerder is opgemerkt dat dit goed te verklaren valt door de wijze waarop de criteriumvariabele zijn geconstrueerd. Het contexteffect is daarmee reeds uitgezuiverd. Het effect van schoolbesturen op de effectiviteit van basisscholen kent tevens tegen de – op een strikte 'inside-out' analyse gebaseerde – verwachting in naast een indirecte ook een directe invloed op de output. Het directe effect van bestuurskenmerken op de schooloutput is in overeenstemming met het causale model van figuur 4. Dat de verwachtingen van figuur 2 niet worden bevestigd, zou het gevolg kunnen zijn van de bij de presentatie van het model van figuur 4 geopperde mogelijkheid dat in het databestand onvoldoende procesvariabelen zijn gemeten, of dat de gemeten procesvariabelen onvoldoende hebben gemeten waar het bij schooleffectiviteit om gaat. Zo zijn er in het geanalyseerde databestand bijvoorbeeld geen gegevens op leraar- en klasniveau.

Een andere verklaring van het directe bestuurseffect is mogelijk vanuit kenmerken van de schoolpopulatie. De gedachtengang hierbij is dat de bestuurseffecten en de effecten van denominatie op de outputmaat in feite effecten zijn van de schoolpopulatie, die in het meetmodel als rechtstreekse bestuurseffecten op de output zichtbaar worden. In verschillende studies waarin reeds op individueel niveau voor leerlingkenmerken is gecorrigeerd wordt dit zogenaamd extra context-effect vastgesteld (vgl. Brandsma, 1993; Knuver, 1993; Hofman, 1993b).

De effecten van denominatie en van de bestuurskenmerken, die zichtbaar werden in de LISREL-analyses, lopen voornamelijk via één van de twee procesvariabelen die in de regressieanalyses als predictor voor rekenprestaties naar voren kwamen. Het betreft evaluatie van leerlingprestaties op schoolniveau. Daarbij gaat het om (a) in hoeverre de school gestandaardiseerde schoolvorderingentoetsten afneemt, (b) in hoeverre deze gebruikt worden om het leerstofaanbod te evalueren en (c) of de leerprestaties centraal worden geregistreerd. Dit resultaat bevestigt de verklaring dat confessionele scholen gemiddeld effectiever zijn ten gevolge van een beter academisch klimaat (Dronkers, 1992).

De veronderstellingen van het tweede en het vierde model voor een causale structuur bleken niet nodig om het meest passend LISREL-model te vinden. Noch de veronderstelling dat denominatie en context een directe invloed hebben op procesvariabelen (figuur 3), noch de veronderstelling dat invloeden van denominatie en context eventueel verlopen via ongemeten procesvariabelen (Figuur 5), zijn nodig geweest.

De resultaten van de secundaire analyses kunnen aanzienlijke praktische consequenties hebben. De resultaten bevestigen dat de bestuursvorm van scholen voor openbaar en bijzonder onderwijs nogal verschilt. Naast de verschillen in schaal en complexiteit is een belangrijk verschilpunt dat de geledingen in en rond de school binnen het openbaar onderwijs weinig invloed uitoefenen op het bestuursbeleid.

Uit onderzoek in de Verenigde Staten naar de wenselijke bestuursvorm voor het openbaar 'public' onderwijs komen soortgelijke resultaten naar voren. Chubb en Moe (1991) stellen vast dat meer dan 60% van de effectieve scholen beschikken over meer dan gemiddelde bestuurlijke autonomie, tegen slechts een derde deel van de niet-effectieve scholen. De bestuursvorm die Chubb en Moe daarbij voorstaan betekent eigenlijk dat het bestuur van de openbare school moet worden losgekoppeld van de publiekrechtelijke overheidssector.

Deze gedachtengang past wonderwel binnen de in Nederland weer opgelaaide discussie omtrent de wenselijke bestuursvorm voor het openbaar onderwijs. Wijzigingsvoorstellen komen meestal neer op een ontkoppeling van de rol van de gemeente als schoolbestuur van de algemene verantwoordelijkheid van de gemeentelijke overheid voor het gehele (lokale) onderwijs. Vele mogelijkheden zijn naar voren gebracht dan wel in de praktijk gerealiseerd, zoals de functionele commissie, de samenwerkingsschool, de privaatrechtelijke stichting of vereniging met het statutair vastgelegd openbaar karakter, de nationale school, het onderwijsschap en dergelijke. Alhoewel de discussie over de wenselijke bestuursvorm in laatste instantie niet langs empirische weg tot een conclusie is te voeren, geven de resultaten van dit onderzoek aanleiding tot een pleidooi voor een bestuursvorm waarbij de geledingen in en rond de school de beslissingen van het bestuur kunnen beïnvloeden; dit betekent dat er een bestuur noodzakelijk is dat dicht bij die ene specifieke school staat.

LITERATUUR

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society Series A (General) 149.* part 1, 1-43.

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

Bosker, R.J., & Hofman, W.H.A. (1987). Dimensies van schoolkwaliteit: de algemene en milieuspecifieke invloed van scholen op de prestaties en het keuzegedrag van leerlingen. In J. Scheerens & W.G.R. Stoel, *Effectiviteit van onderwijsorganisaties.* Lisse: Swets & Zeitlinger.

Brandsma, H.P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs.* Proefschrift. (Monografieën onderwijsonderzoek 14). Groningen: RION.

Braster, J.F.A., & Leune, J.H.G. (1986). *Samenstelling en functioneren van schoolbesturen.* Rotterdam: Erasmus Universiteit, Vakgroep Onderwijssociologie en Onderwijsbeleid.

Buurke, W.J. (1988). *Het gemeentebestuur als lokale overheid en als schoolbestuur.* Nijmegen: ITS.

Chubb, J.E., & Moe, T.M. (1991). *Politics, markets and America's schools.* Washington D.C.: The Brookings Institutions.

Creemers, B.P.M., & Scheerens, J. (1991). Onderwijseffectiviteit: overwegingen voor een programma van onderzoek. *Tijdschrift voor Onderwijsresearch, 16* (4), 193-206.

Detering, P., & Kalkman, W. (1985). *Besturen met beleid.* Amsterdam: Universiteit van Amsterdam, Vakgroep Onderwijskunde.

Dronkers, J. (1992). Blijvende organisatorische onderwijsverzuiling ondanks secularisering: een onbedoeld effect van overheidsbeleid? *Beleid & Maatschappij,* 5, 227-237.

Esch, W. van (1988). *Schoolbesturen en de vrijheid van onderwijs.* Dissertatie. Nijmegen: ITS.

Eeden, P. van, & Meijnen, G.W. (1990). Themanummer Multi-niveau onderzoek: uitgangspunten en toepassingen. *Tijdschrift voor Onderwijsresearch, 15,* (5), 259-333.

Hofman, R.H. (1993a). *Effectief schoolbestuur, Een studie naar de bijdrage van school besturen aan de effectiviteit van basisscholen.* Proefschrift. (Monografieën onderwijsonderzoek 13). Groningen: RION.

Hofman, W.H.A. (1993b). Effectief onderwijs aan allochtone leerlingen. Proefschrift Delft: Eburon.

Hofman, R.H., & Hofman, W.H.A. (1988). Definiëring van onderwijseffectiviteit: kanttekeningen vanuit het onderwijsonderzoek. In B.P.M. Creemers (Red.), *Handboek voor Schoolorganisatie en Onderwijsmanagement.* Alphen aan den Rijn: Samson, 6310-1 t/m 6310-17.

Knuver, J.W.M. (1993). *De relatie tussen klas- en schoolkenmerken en het affectief functioneren van leerlingen.* Proefschrift. (Monografieën onderwijsonderzoek 15). Groningen: RION.

Longford, N.T. (1986). *Statistical modelling of data from hierarchical structures using variance component analysis.* Lancaster University: Centre for Applied Statistics.

Madaus, G.F. (1980). *School effectiveness: a reassessment of the evidence.* New York: Mc Graw-Hill.
Meijnen, G.W. (1984). *Van zes tot twaalf.* Den Haag: Stichting voor Onderzoek van het Onderwijs (SVO).
Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *The junior school project: technical appendices.* London: ILEA, Research and Statistics Branch.
Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement, 1,* (1), 61-80.

Manuscript ontvangen 28-7-1993
Definitieve versie ontvangenm16-8-1994

BIJLAGE I

Psychometrische karakteristieken van contingentiefactoren, structuurkenmerken en schoolprocesvariabelen.

Tabel 1. Psychometrische karakteristieken van drie contingentiefactoren (gemiddelde, standaard deviatie, aantal items, bereik en Cronbach's alpha).

	X	(SD)	items	range	gest. α
(1) **Schaalgrootte en complexiteit**:					
(a) Aantal basisscholen	6.7	(8.7)	1	1-44	nvt
(b) Regionale positie	1.9	(1.3)	1	0- 4	nvt
(c) Aantal schooltypen	1.2	(2.5)	3	0- 4	.80
(2) **Omgevingsinvloed:**					
(a) Aantal bestuursvergaderingen met reguliere geledingen	6.2	(2.3)	5	1-10	.80
(b) Invloed van geledingen	2.0	(0.5)	8	1- 4	.73
(c) Invloed van externe groeperingen	1.9	(0.5)	13	1- 4	.77
(3) **Externe ondersteuning:**					
(a) Aantal instellingen externe ondersteuning	0.9	(0.3)	4	0-13	.70
(b) Breedte van externe ondersteuning	2.0	(0.5)	5	0- 5	.73

Tabel 2. Structuurkenmerken van schoolbesturen.

Structuurkenmerken	X	(sd)	items	range	gest. α
Bestuursdifferentiatie					
(1) Interne bestuursopbouw	3.7	(1.1)	3	2-6	.80
(2) Aantal interne organen	2.2	(3.7)	3	0-29	som
Deskundigheid en cohesie					
(3) Deskundigheid	3.2	(0.5)	4	1-4	.68
(4) Vakdeskundigheid	2.1	(1.7)	5	0-5	.77
(5) Cohesie	3.4	(0.6)	1	1-4	nvt
Planmatigheid					
(6) Planmatigheid besluitvorming	11.6	(1.0)	3	4-12	.72
(7) Planmatigheid beleid	8.2	(1.5)	5	2-10	.62

Tabel 3. Resultaten van schaalconstructie taakgebieden.

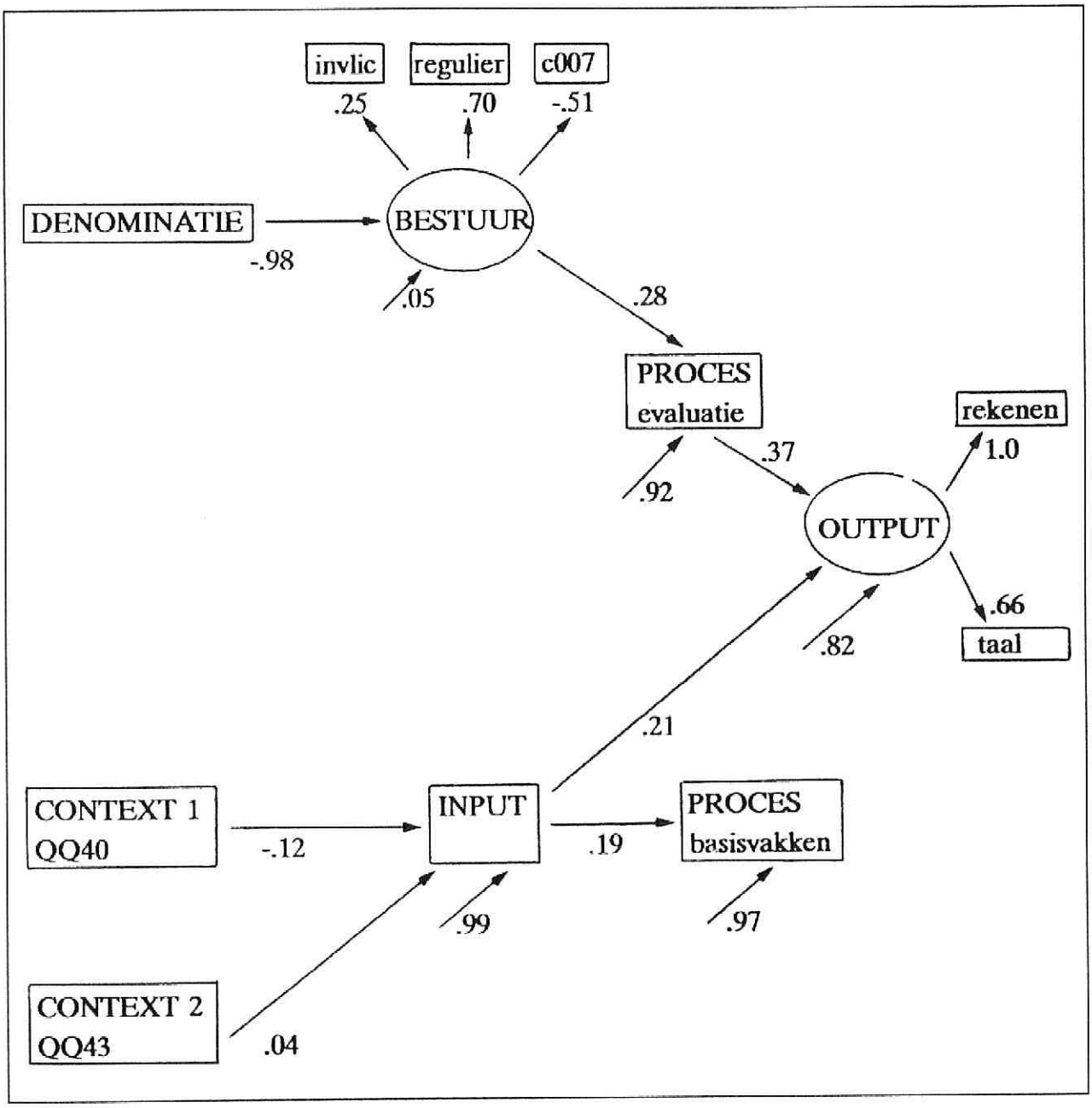
Schalen per taakgebied	X	(sd)	items	range	gest. α
(1) Interne contacten					
ICFREQ	2.2	(0.5)	8	1-5	.72
ICTAAK	1.5	(0.6)	8	1-4	.52
(2) Externe contacten					
ECFREQ	2.0	(0.5)	12	1-5	.75
ECTAAK	1.6	(0.6)	12	1-4	.69
(3) Onderwijsbeleid					
OBFREQ	1.7	(0.4)	16	1-5	.86
OBTAAK	1.4	(0.7)	16	1-4	.81
(4) Organisatiebeleid					
ORFREQ	1.7	(0.4)	18	1-5	.86
ORTAAK	1.3	(0.6)	18	1-4	.80
(5) Personeelsbeleid					
PRFREQ	1.9	(0.5)	11	1-5	.76
PRTAAK	1.2	(0.6)	11	1-4	.67
(6) Financieel/materieel beleid					
FMFREQ	2.1	(0.6)	13	1-5	.83
FMTAAK	1.8	(0.8)	13	1-4	.82

(Toelichting: X = gemiddelde somscore van alle items van een schaal; α = Cronbachs alpha-coëfficiënt voor interne consistentie).

Tabel 4. Psychometrische kenmerken van schoolprocesvariabelen.

SCHOOL PROCESVARIABELEN	M	(sd)	items	range	gest. α/H
1. onderwijskundig leiderschap	2.2	(.48)	11	1-4	.81
2. prestatie-oriëntatie	2.5	(.63)	2	2-4	nvt
3. evaluatie van onderwijs op schoolniveau	4.0	(1-6)	3	1-6	nvt
4. nadruk op basisvaardigheden	3.1	(.70)	6	2-5	H=.53
5. (a) mate van orde en	4.5	(.41)	11	3-5	.83
(b) mate van regelgeving	2.0	(.69)	4	1-4	H=.67
6. evaluatie-gerichtheid leerlingbeleid	4.2	(1.31)	3	1-6	nvt

BIJLAGE II



Chi^2=104.29, df=40, p=.0000

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, pp. 79-92

Het ontwerpen van tests met verschillende optimaliteitscriteria

Martijn P.F. Berger en Wim J. van der Linden*

ABSTRACT

The problem of finding optimal designs has been extensively studied in the statistical literature. Optimal design procedures and criteria can also be used for the assembly of psychological and achievement tests. In this paper the problem of designing optimal tests will be considered within the context of item response theory. Optimal test designs based on different criteria will be considered and compared with each other. The results and conclusions will be useful for test constructors.

INLEIDING

In sociaal-wetenschappelijk onderzoek worden op verschillende momenten onderzoeksontwerpen gebruikt. Het ontwerp van een experiment wordt vaak met het woord '*proefopzet*' of met de Engelse term '*design*' aangeduid. Een experimenteel design wordt bepaald door de combinaties van de experimentele condities of onafhankelijke variabelen en de te meten afhankelijke variabele. Hoewel de termen '*proefopzet*' en '*design*' het eerst in experimenteel onderzoek werden gebruikt, zijn deze begrippen langzamerhand ook in andere vormen van onderzoek ingeburgerd. Zo wordt bijvoorbeeld in survey-onderzoek van een '*onvolledig design*' gesproken, om aan te geven dat bepaalde steekproefgegevens ontbreken.

Hoewel meestal inhoudelijke overwegingen bepalend zijn voor de keuze van een onderzoeksontwerp, wordt het ontwerp mede bepaald door het gehanteerde statistische model en de hierin gedefinieerde parameters. Wanneer bijvoorbeeld verondersteld wordt dat een bepaald variantie-analytisch model de relatie tussen de onafhankelijke variabelen en de afhankelijke variabele adequaat beschrijft, dan zou een design waarbij de effectparameters zo efficiënt mogelijk geschat kunnen worden, d.w.z. met een zo klein mogelijke variantie voor de schatter, de beste keuze zijn. Het is uit experimenteel onderzoek bekend dat het variëren van de experimentele condities de variantie van de schattingen van de parameters kan beïnvloeden. Vaak worden door onderzoekers de condities van de experimentele proefopzet op zodanige wijze gekozen, dat het experiment een zo klein mogelijke variantie voor de parameterschatters oplevert. Door deze keuze kan men de benodigde hoeveelheid experimentele gegevens aanzienlijk reduceren.

In de statistische literatuur is uitgebreid onderzoek gedaan naar het optimaliseren van experimentele proefopzetten. De statistische theorie waarop dit onderzoek gebaseerd is wordt meestal aangeduid met de term '*optimal design theory*'. Een overzicht van de ontwikkelingen op het gebied van optimaal design onderzoek wordt o.a. gegeven door Atkinson (1982), Silvey (1980) en Steinberg en Hunter (1984). In het algemeen kan gesteld worden dat het optimaal ontwerpen van een onderzoek enerzijds afhangt van het veronderstelde model met de bijbehorende parameters en anderzijds van het gekozen optimaliteitscriterium. Zo zal een optimaal onderzoeksontwerp voor een lineair (variantie-analytisch of regressie-) model dus niet automatisch optimaal zijn voor een niet-lineair (polynoom of exponentieel) model. Ook het criterium op grond waarvan de optimaliteit wordt vastgelegd, zal bepalend zijn voor de keuze van een optimaal design. In het vervolg van dit artikel zal met name op dit laatste punt nader worden ingegaan.

---

* De auteurs staan in alphabetische volgorde vermeld en zijn in gelijke mate verantwoordelijk voor de inhoud van dit artikel.
Adres: Faculteit der Toegepaste Onderwijskunde, Universiteit Twente, Postbus 217, 7500 AE Enschede.

Recentelijk zijn de resultaten van onderzoek uit de statistische literatuur naar optimale designs aangewend voor het optimaal verzamelen van gegevens voor het schatten van test- of populatieparameters in peilingsonderzoek. Lord (1962) en Pandey en Carlson (1976) hebben met behulp van '*multiple-matrix sampling*' aangetoond dat binnen de context van de klassieke testtheorie de precisie van de schatting van de gemiddelde testscore van een domein het nauwkeurigste is, wanneer ieder item door een andere groep personen wordt beantwoord. Met behulp van 'multiple-matrix designs' werd het mogelijk om het aantal items dat aan leerlingen voorgelegd moet worden om een bepaalde precisie voor de schattingen te verkrijgen, aanzienlijk te reduceren. Dergelijke reducties in afnametijd zullen vooral van belang zijn bij grootschalige peilingsprojecten. Een voorbeeld van een dergelijk project is PPON (Verhelst & Eggen, 1989).

De belangrijkste reden voor het afnemen van psychologische tests en studietoetsen is om de latente vaardigheden van personen zo nauwkeurig mogelijk te schatten. Omdat testgegevens in toenemende mate beschreven worden met modellen uit de itemresponstheorie (IRT), is het vaststellen van optimale designs voor testgegevens ook voor deze modellen belangrijk. In dit artikel zal voornamelijk ingegaan worden op de manier waarop tests aan de hand van een viertal optimaliteitscriteria ontworpen of samengesteld kunnen worden om efficiënte vaardigheidschattingen voor personen onder een IRT model te verkrijgen. Het model dat als voorbeeld gekozen wordt, is telkens het twee-parameter logistische model, dat hieronder nader gedefinieerd zal worden. Op de feitelijk te gebruiken algoritmen om optimale tests te verkrijgen zal niet nader worden ingegaan. Voor een beknopt overzicht van het gebruik van mathematische programmering bij test selectie wordt verwezen naar Theunissen, Sanders & Verschoor (1993).

In het vervolg wordt aangenomen dat een (oneindig grote) gecalibreerde itembank beschikbaar is, waarvan de itemparameters dus bekend zijn. De vraagstelling is hoe we met de items uit deze bank een optimale test voor de meting van de vaardigheden in een populatie personen (schoolklas, jaargroep) kunnen ontwerpen. In eerste instantie is de informatie die we onder deze aanname verkrijgen alleen theoretisch van aard, omdat de toetsconstructeur in de praktijk zo'n ideale itembank niet ter beschikking heeft. Uit deze informatie kan evenwel een aantal principes voor testdesign gedestilleerd worden die praktische betekenis hebben en waarmee niet alleen professionele testconstructeurs, maar ook leraren en anderen die eigen tests samenstellen hun voordeel mee kunnen doen. Eerst zal echter aan de hand van het twee-parameter logistische model aangegeven worden hoe de methodologie voor het optimaal samenstellen van een test werkt en daarna zullen aan de hand van enkele analyses een aantal conclusies geformuleerd worden.

OPTIMALE TESTDESIGN

In IRT modellen wordt de kans dat een persoon j een bepaalde respons geeft op een item i weergegeven als een functie van de vaardigheid θ_j van die persoon en van één of meer parameters ξ_i, welke item i karakteriseren. Een overzicht van verschillende IRT modellen is te vinden in Hambleton and Swaminathan (1985, hdst. 3). Het twee-parameter logistische IRT model is bijvoorbeeld als volgt weer te geven:

$$P\,\theta_j \mid \xi_i) = \{1 + \exp\,[-\xi_{2,i}(\theta_j - \xi_{1,i}]\}^{-1}. \qquad (1)$$

In (1) is te zien dat dit model gekenmerkt wordt door slechts twee itemparameters, n.l. $\xi_i = (\xi_{1,i}, \xi_{2,i})$, waar $\xi_{1,i}$ de moeilijkheidsparameter van het item i voorstelt en $\xi_{2,i}$ de discriminatieparameter is.

Wanneer nu een IRT model gebruikt wordt, dan worden eerst de itemparameters voor een verzameling van items uit een steekproef met responsgegevens geschat en wordt de passing van het model getoetst. Na deze fase van itemcalibratie en modeltoetsing, is het model klaar om als meetmodel gebruikt te worden. Iedere willekeurige deelverzameling items uit een gecalibreerde itembank kan als test bij een persoon afgenomen worden en uit de aldus verkregen responsen worden dan de vaardigheidsparameters θ_j van de personen geschat.

Verschillende soorten testdesigns zijn mogelijk. Stel dat de test uit n items moet bestaan en bij een groep van N personen afgenomen moet worden. Als ieder van deze N personen alle n items voorgelegd krijgt, dan is het resultaat een Nxn matrix die volledig gevuld is met Nxn responsen. Een voorbeeld van dit design, dat in de praktijk het meest voorkomt, wordt gegeven in de eerste matrix in Figuur 1. Een tweede soort design ontstaat als deelverzamelingen uit een grotere verzameling van n items ieder aan andere deelverzamelingen van N personen voorgelegd worden. De Nxn design matrix met responsen is dan onvolledig. Onvolledige designs van dit type komen bijvoorbeeld voor in peilingsonderzoek. Een voorbeeld van een onvolledig design is weergegeven in de tweede matrix in Figuur 1. Het laatste soort design is eveneens onvolledig en ontstaat wanneer een toets voor ieder van de N personen adaptief uit de verzameling van n items samengesteld wordt. Een voorbeeld van een adaptief design is in de derde matrix in Figuur 1 weergegeven. In dit voorbeeld blijkt dat elke persoon een andere verzameling van items te beantwoorden kan krijgen. Hoewel de laatste twee designs onvolledig zijn, bestaat er een belangrijk verschil tussen beide. In de tweede matrix in Figuur 1 zijn de afwezige responsen structureel, d.w.z. vooraf bepaald door de testconstructeur. In het adaptieve design in de derde matrix van Figuur 1 hangt het al of niet voorkomen van een respons af van de voorafgaande responsen. Dit design is dus probabilistisch van aard.

Definitie van het begrip testdesign

We vervangen het voorafgaande, terloops gebruik van de term '*design*' door een nauwkeurige omschrijving in termen van de parameters in een IRT model, zoals bijvoorbeeld het tweeparameter logistische model (1). Hoewel IRT modellen uitgaan van een vaardigheidsverdeling die continue is, wordt in dit artikel uit practische overwegingen gekozen voor een discrete vaardigheidsschaal. Dit betekent dat de vaardigheidsparameter θ in c discrete punten θ_j beschouwd wordt, zodat $\theta = (\theta_1, \ldots, \theta_c)$. De keuze van het aantal discrete vaardigheidspunten c en van de afstanden tussen deze punten is in principe vrij, maar deze dienen uiteraard wel zodanig gekozen worden dat de beoogde vaardigheidsverdeling zo goed mogelijk benaderd wordt. Verder definiëren we een vector met gewichten $W = (w_1, \ldots, w_c)$, waarvan ieder gewicht aangeeft hoe belangrijk we het corresponderende vaardigheidsniveau θ_j vinden. Deze gewichten kunnen verschillende invullingen krijgen (Veerkamp & Berger, 1993). In de psychometrische literatuur staat deze vector ook wel bekend als de '*target information function*' of doelinformatiefunctie (Hambleton & Swaminathan, 1985). Een mogelijke interpretatie van deze gewichten is dat de vector W de voorkennis weergeeft die de testconstructeur heeft over de vaardigheidsverdeling van de personen. Gelijke gewichten wijzen dan bijvoorbeeld op de aanname van een uniforme vaardigheidsverdeling. Het idee bij deze interpretatie van W is dat een vaardigheidsniveau meer

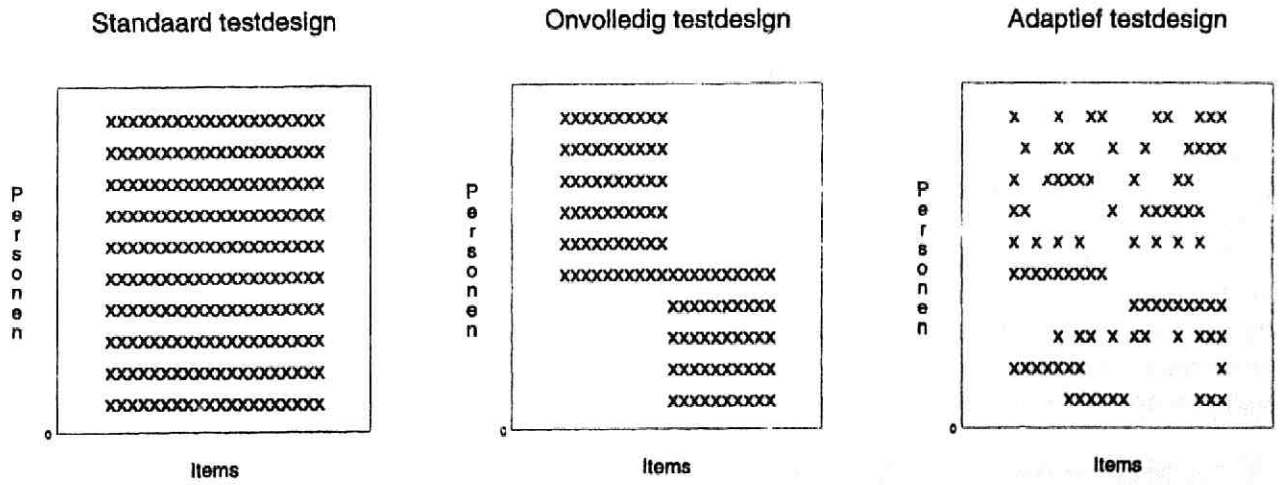


Fig. 1. Drie typische testdesigns.

gewicht krijgt in het ontwerp naarmate het vaker in de doelgroep personen voorkomt. Er zijn evenwel ook andere redenen om bepaalde gewichten te kiezen. Wanneer bijvoorbeeld een bepaalde grensscore in een selectieprocedure wordt vastgesteld, dan moet de overeenkomstige vaardigheid een relatief hoog gewicht krijgen. Voor het gemak zullen we in het vervolg evenwel over het paar $\{\theta, W\}$ spreken als de a priori (discrete) vaardigheidsverdeling van de groep personen. De itemparameters die een test met n items karakteriseren zullen door de vector $\xi = (\xi_1, \ldots, \xi_n)$ worden weergegeven.

Een testdesign wordt nu gerepresenteerd door een z.g. designvector $p = (p_1, \ldots, p_n)$, waarvan de elementen p_i aangeven hoe vaak ieder van de parameterwaarden $\xi = (\xi_1, \ldots, \xi_n)$ in de test voor moet komen. Afhankelijk van het veronderstelde IRT model kunnen de elementen ξ_i scalaire grootheden of vectoren zijn met een aantal itemparameters. Zoals al aangegeven werd, is voor het 2-parameter logistische model in (1) iedere component van ξ een vector met een moeilijkheid- en een discriminatieparameter. In dit artikel wordt aangenomen dat de elementen van de vector p proporties of percentages zijn.

OPTIMALITEITSCRITERIA

Het vinden van een optimaal testdesign kan in het algemeen gekarakteriseerd worden als optimalisatie onder randvoorwaarden. De grootheid die geoptimaliseerd wordt, kan bijvoorbeeld zijn: de betrouwbaarheid van de test, de standaardmeetfout van de test in een bepaald punt, de overeenstemming tussen de test en een bepaalde ideale blauwdruk voor de inhoud van de test of de benodigde afnametijd van de test. In dit artikel zal de methodologie van optimaal testdesign besproken worden met als doel de maximalisering van de efficiëntie van de schatters van θ_j. Optimaliteitscriteria voor dit doel zullen eerst uitgewerkt worden. Daarna wordt teruggekomen op de randvoorwaarden die bij het optimaliseren een rol kunnen spelen.

Als maat voor de efficiëntie van de schatter van θ_j kan Fisher's informatiemaat, $I(\theta_j|\xi)$, dienen. Een definitie van deze maat wordt gegeven door Kendall en Stuart (1973, p. 10). Voor bekende itemparameters geldt dat deze maat gelijk is aan de inverse van de asymptotische variantie van de schatter van θ_j. Voor het twee-parameter logistische model kunnen we daarom stellen:

$$I(\theta_j|\xi) \equiv \{Var(\theta_j|\xi)\}^{-1} = \sum_{i=1}^{n} \xi_{2,i}^2 P(\theta_j|\xi_i)\,[1-P(\theta)_j|\xi_i)]. \tag{2}$$

De informatie van een testdesign voor de vaardigheidsverdeling $\{\theta, W\}$ kan nu weergegeven worden door de vector:

$$I(\theta|\xi) = [I(\theta_1|\xi), I(\theta_2|\xi), I(\theta_3|\xi), \ldots, I(\theta_c|\xi)]. \tag{3}$$

en de efficiëntie van een design is weer te geven als een functie van deze vector $I(\theta|\xi)$ en de vector W. In het algemene geval van de simultane schatting van meerdere parameters moet een generalisering van (2) gebruikt worden die bekend staat als Fisher's informatiematrix. Omdat we evenwel aan kunnen nemen dat de personen de toets onafhankelijk van elkaar zullen maken, bevatten de covarianties tussen responsen geen informatie over de vaardigheden en is alleen de diagonaal van deze matrix relevant. De vector in (3) is gelijk aan de diagonaal van Fisher's informatiematrix.

Een optimaal testdesign wordt nu gevonden door voor een bepaalde vector $\xi = (\xi_1, \ldots, \xi_n)$ de vector $p = (p_1, \ldots, p_n)$ zodanig te kiezen dat de efficiëntie van het design zo groot mogelijk en dus de variantie van de schatters van θ zo klein mogelijk wordt. Dit kan gerealiseerd worden door een bepaalde functie $\Phi(.)$ van (3) en de vector W te maximaliseren. Het probleem van het vinden van een optimale verzameling items met parameters ξ kan dus als volgt worden geformuleerd:

Maximaliseer $\Phi\{I(\theta|\xi), W\}$ (4)
onder bepaalde voorwaarden.

In de statistische literatuur over optimale designs worden diverse functies of criteria $\Phi(.)$ voorgesteld. In de volgende secties wordt nader op de keuze van het criterium $\Phi(.)$ ingegaan.

D-optimaliteit

Het meest bekende criterium is het produkt- of D-optimaliteitscriterium:

$$\Phi \{I(\theta|\xi), W\} = \prod_{j=1}^{c} I(\theta_j|\xi)^{w_j}. \tag{5}$$

Dit criterium wordt ook wel het determinantcriterium genoemd, wanneer de functie gebaseerd is op een matrix in plaats van een vector, dus in het geval waarin de covarianties tussen de itemresponsen niet verwaarloosd mogen worden. Er zijn een aantal redenen te noemen waarom dit criterium zoveel gebruikt is bij het optimaliseren van een design. Een belangrijke eigenschap van het D-optimaliteitscriterium is dat het invariant is onder lineaire transformaties van de parameters. Bovendien zijn voor speciale gevallen bovengrenzen van het criterium vast te stellen (zie Khan & Yazdi, 1988). Een meer uitgebreide beschrijving van de voor- en nadelen van dit criterium wordt gegeven door Berger (1992b) en Berger en van der Linden (1992).

Een veel gebruikte procedure om D-optimale designs te verkrijgen is een sequentiële procedure. Een stapsgewijze selectie van items vindt plaats door telkens uit alle beschikbare items in een itembank het item (k+1) te selecteren dat de maximale waarde heeft voor:

$$\Phi \{ [I(\theta|\xi^{(k)}) + I(\theta|\xi^{(k+1)})] , W\}, \tag{6}$$

waarbij $\xi^{(k)}$ een vector is met parameters van alle items die al eerder geselecteerd zijn.
Nadere details van deze procedure zijn te vinden in Berger (1992a, 1994).

A-optimaliteit

Naast het D-optimaliteitscriterium wordt het som – of A-optimaliteitscriterium ook veel gebruikt. Dit criterium is gedefinieerd als:

$$\Phi \{I(\theta|\xi), W\} = \sum_{j=1}^{c} w_j I(\theta_j|\xi). \tag{7}$$

Hoewel het criterium intuïtief aantrekkelijk is, heeft het niet de hierboven gegeven voordelen van het D-criterium. Een belangrijk verschil tussen beide criteria is dat het A-criterium sterkere compensatie tussen de informatie in de steekproef over verschillende parameterwaarden θ_j in de vector met vaardigheidsparameters toelaat dan het D-criterium. Enigszins generaliserend zou gesteld kunnen worden dat het D-criterium conjunctief van aard is, terwijl het A-criterium disjunctief is.

Voor de berekening van een A-optimaal design kan gebruik gemaakt worden van een eenvoudig lineair programmeringsmodel (LP-model). De berekeningen verlopen als volgt. Laat p_i een willekeurig element in de designvector p zijn, waarbij i=1, ..., n. Deze elementen zijn nu de (reëel-waardige) beslissingsvariabelen in het volgende LP-model:

$$\text{Maximaliseer} \sum_{j=1}^{c} \sum_{i=1}^{n} w_j I(\theta_j|\xi_i) p_i \tag{8}$$

met als randvoorwaarden

$$\sum_{i=1}^{n} p_i = 1, \tag{9}$$

en

$$p_i \geq 0 \quad i=1,...,n. \tag{10}$$

De optimale waarden van de vector p kunnen voor dit model uitgerekend worden met behulp van standaard LP-programmatuur.

E-optimaliteit

Het volgende criterium dat hier besproken wordt is het E-optimaliteitscriterium. Dit criterium is gedefinieerd als:

$$\Phi \{I(\theta|\xi), W\} = \min_j [I(\theta_j|\xi)]. \tag{11}$$

In dit geval wordt van alle parameters in de vector θ de parameter met de kleinste informatie in het criterium opgenomen. Men zou dit nader kunnen omschrijven als de situatie waarbij men het zekere voor het onzekere neemt, namelijk de minst informatieve parameter θ wordt gebruikt om een optimale test samen te stellen. Merk op dat dit criterium in feite onafhankelijk is van W en dat dit criterium daarom minder geschikt is om een optimale test samen te stellen (d.w.z. conservatiever wordt) als de verdeling van de vaardigheid verder afwijkt van een uniforme verdeling. Een meer algemene behandeling van dit criterium is te vinden in Van der Linden (1994).

Ook voor de berekening van een E-optimaal design kan LP gebruikt worden. Het model voor dit criterium luidt:

$$\text{Maximaliseer} \quad y \tag{12}$$

onder de randvoorwaarden

$$\sum_{i=1}^{n} I(\theta_j|\xi_i)p_i \geq y, \quad \text{voor } j=1,...,c, \tag{13}$$

$$\sum_{i=1}^{n} p_i = 1, \tag{14}$$

$$p_i \geq 0, \quad i=1,...,n. \tag{15}$$

In het model definiëren (12) en (13) de beslissingsvariabele y als de waarde van de informatiefunctie die het minimum vormt over de vector met parameterwaarden θ. In (12) wordt deze variabele gemaximaliseerd.

Gewogen E-optimaliteit

Het laatste criterium dat hier geïntroduceerd wordt, kan opgevat worden als een gewogen versie van het E-optimaliteitscriterium. Dit criterium ontstaat als we in het model in (12)-(15) de beslissingsvariabele y van de gewichten w_i voorzien, dus als we de randvoorwaarde in (13) vervangen door:

$$\sum_{i=1}^{n} I(\theta_j|\xi_i) \geq w_i y, \quad \text{voor } j=1,...,c. \tag{16}$$

Het resultaat is dat de informatie over alle parameters in θ gemaximaliseerd wordt, maar dat daarbij de gewichten in de vector W als randvoorwaarden opgelegd worden. Omdat het om een maximaliseringsprobleem gaat, zal de informatieverdeling over de parameters daarom de waarden van de gewichten volgen. Merk op dat door de opname van de gewichten in het model het gewogen E-criterium in tegenstelling tot het ongewogen E-criterium niet meer conservatief is. Het gewogen E-optimaliteitscriterium werd geïntroduceerd onder de naam *'maximin model for test design'* in Van der Linden & Boekkooi-Timminga (1989).

Overige voorwaarden bij optimalisatie

In de praktijk worden naast de randvoorwaarden die in (8)-(14) vermeld staan diverse andere testspecificaties als randvoorwaarde aan de optimalisering opgelegd. Een bekend onderscheid in de testconstructie is dat tussen betrouwbaarheid en validiteit. Enigszins vereenvoudigend kan gesteld worden dat de doelfunctie in (4) garandeert dat het testontwerp maximale betrouwbaarheid krijgt, terwijl alle aspecten van de test die met de (inhouds)validiteit van doen hebben als

randvoorwaarden meegenomen moeten worden. Drie voorbeelden van klassen van mogelijke randvoorwaarden worden hier verder toegelicht:

1. De eerste klasse van voorwaarden richt zich op de inhoud van de test. Hierbij kan gedacht worden aan voorwaarden met betrekking tot de classificatie van de items volgens van toepassing zijnde inhoudsdimensies, gedragsniveau's, het format van de items of de hoofdstukken in een leerboek. De beide eerste classificaties worden vaak gecombineerd in de vorm van een taxonomie of toetsmatrijs die aan de inhoud van de test opgelegd moet worden. Bekende gedragsniveau's zijn het reproduceren van feiten, het beantwoorden van kennisvragen of het toepassen van regels. Bij classificatie volgens itemformat moet gedacht worden aan vraagvormen als meerkeuze, invullen, goed-fout, essay, e.d. Voor al deze indelingen geldt dat ze zowel afzonderlijk als in combinatie een partitie van de itembank opleveren. De randvoorwaarden in deze klasse nemen de vorm aan van voorschriften met betrekking tot de verdeling van de items in de test over deze partities.
2. Een geheel andere klasse van voorwaarden maakt gebruik van (kwantitatieve) parameters die de eigenschappen van de items beschrijven. Enkele voorbeelden hiervan zijn de eis dat de testlengte beperkt moet blijven tot een bepaald aantal items, de eis dat de itemmoeilijkheden alle in een bepaald interval moeten liggen, of de eis dat de items minder dan een bepaald aantal malen eerder gebruikt mogen zijn. De randvoorwaarden in deze klasse kunnen alle geformuleerd worden als voorschrift dat de individuele waarden van de (kwantitatieve) parameters voor de items in een bepaald interval moeten liggen of dat dit geldt voor statistieken als bijvoorbeeld hun som.
3. De laatste klasse van randvoorwaarden die hier onderscheiden wordt, betreft de relaties tussen de items in de itembank. Voorbeelden zijn de relaties van insluiting en uitsluiting. Het eerste voorbeeld doet zich voor als sommige items zich op een gemeenschappelijke tekst of opgave richten en de items alleen als blok wel of niet voor de test geselecteerd mogen worden. Items kunnen elkaar daarentegen uitsluiten, wanneer een item een hint bevat in de richting van de oplossing van een ander item. Deze twee items mogen dan nooit tegelijkertijd voor dezelfde test geselecteerd worden. In een voorbeeld dat verderop gegeven wordt, zal aangegeven worden hoe een optimale oplossing onder dergelijke voorwaarden berekend kan worden.

Een volledige behandeling van de randvoorwaarden die in de praktijk voor kunnen komen, wordt hier niet nagestreefd. Een compleet overzicht van randvoorwaarden in de vorm van lineaire (on)gelijkheden die in LP-modellen voor optimaal testontwerp opgenomen kunnen worden, is te vinden in Van der Linden en Boekkooi-Timminga (1989).

ENKELE VOORBEELDEN VAN OPTIMALE TESTDESIGNS

Om inzicht te geven in de wijze waarop een optimaal testdesign vastgesteld kan worden, worden in deze paragraaf enkele hypothetische situaties geschetst.

Stel dat een testconstructeur een toets wil samenstellen die een bepaalde vaardigheid meet van een populatie leerlingen. Het is niet precies bekend hoe de vaardigheid in deze populatie verdeeld is. De constructeur heeft evenwel goede gronden om te veronderstellen dat de vaardigheidsverdeling bij benadering normaal is, maar zou ook niet vreemd opkijken als deze enigszins positief of negatief scheef is. Wanneer uitgegaan wordt van een vaardigheidsschaal die loopt van $\theta_1 = -3$ naar $\theta_C = +3$, dan kunnen de gewichten W zodanig gekozen worden dat deze een normale verdeling benaderen dan wel een positief of negatief scheve verdeling.

In de Figuren 2-7 zijn (hypothetische) optimaal samengestelde tests voor de vier geformuleerde optimaliteitscriteria weergegeven. De berekeningen vonden plaats voor het twee-parameter logistische model in (2), met behulp van de eerder geschetste algoritmen. Voor het D-optimaliteitscriterium werd een eigen geschreven computerprogramma gebruikt, terwijl voor de overige criteria het LP-programma PCPROG (Kalvelagen, 1986) werd gebruikt. In deze figuren

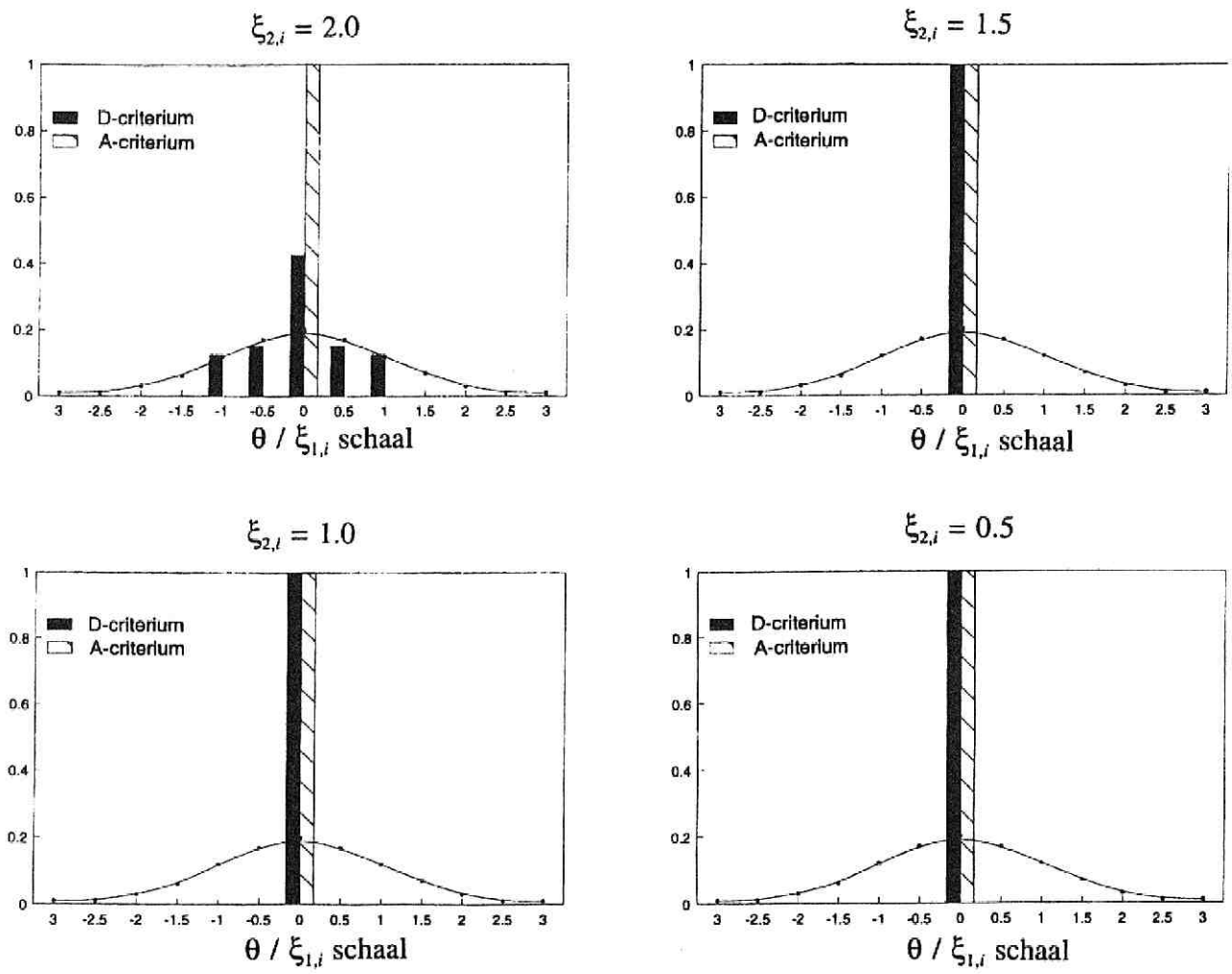


Fig. 2. Optimale tests voor een normale vaardigheidsverdeling samengesteld met het D- en A- optimaliteitscriterium.

worden de relatieve frequenties van een normale en een positief scheve vaardigheidsverdeling weergegeven. Ter illustratie zijn eveneens de resultaten voor een uniforme verdeling toegevoegd. De histogrammen geven de proporties van het aantal items in de optimaal samengestelde test weer, telkens voor een bepaalde combinatie van waarden van de beide itemparameters. Hierbij wordt de itemmoeilijkheid gesymboliseerd door ($\xi_{1,i}$) en de discriminatieparameter door ($\xi_{2,i}$).

In de Figuren 2-4 zijn de resultaten voor het D- en A-optimaliteitscriterium samengenomen. Deze resultaten laten de volgende conclusies toe:

1. Als de waarden voor de discriminatieparameters laag zijn (d.w.z. niet groter dan 0,5) is het voor beide criteria de verstandigste strategie om items te kiezen die alle een waarde voor de moeilijkheidsparameter in het centrum van de vaardigheidsverdeling bezitten. De resultaten voor de scheve verdeling suggereren dat de beste maat voor het centrum van de verdeling niet de modus, maar de mediaan of het gemiddelde is.
2. Voor het A-criterium geldt dit principe voor *alle* bestudeerde waarden van de discriminatieparameter.
3. Voor het D-criterium geldt dat bij toename van het discriminerend vermogen van de items de waarden voor de moeilijkheidsparameter de tendens vertonen om de vorm van de vaardigheidsverdeling te volgen. Deze tendens is sterker naarmate de vaardigheidsverdeling meer de uniforme verdeling benadert.

Interessant is het gedrag van de optimale waarden voor de moeilijkheidsparameters voor de uniforme vaardigheidsverdeling. Deze vertoont de neiging om met toenemende waarden voor de discriminatieparameter van ééntoppig naar tweetoppig over te gaan. Een dergelijk effect werd ook al door Stocking (1990) en door Berger (1992b) opgemerkt bij het optimaal samenstellen van steekproeven voor de schatting van itemparameters. Dit verschijnsel is voor een belangrijk gedeelte het gevolg van het feit dat (1) de vaardigheidsschaal aan de linker- en rechterkant afgeknot is; (2) er relatief veel informatie aan de eindpunten van het interval gevraagd wordt; terwijl (3) de iteminformatiefuncties voor het 2-parameter logistische model symmetrisch zijn. Bij zwakke discriminatie zijn de iteminformatiefuncties relatief vlak en is het verstandig om alle items midden op de schaal te kiezen, zodat geen informatie verloren gaat doordat deze buiten het interval valt. Als de discriminatie toeneemt, worden de iteminformatiefuncties smaller en gepiekter. Er moeten dan items naar het einde van het interval opschuiven, maar niet volledig naar het einde omdat dan weer informatie verloren zou gaan. Door hun symmetrie leveren de informatiefuncties van deze items ook informatie naar het midden van de schaal op, waardoor de lege plekken naast de twee grote pieken in de grafiek voor $\xi_{2,i} = 2.0$ in Figuur 4 zijn ontstaan. Het is daarom van belang om te weten dat zowel het gekozen vaardigheidsinterval als de waarden voor de discriminatieparameters in de figuren realistische zijn en veelvuldig voorkomen in empirische analyses met een standaardnormering van de itemmoeilijkheden.

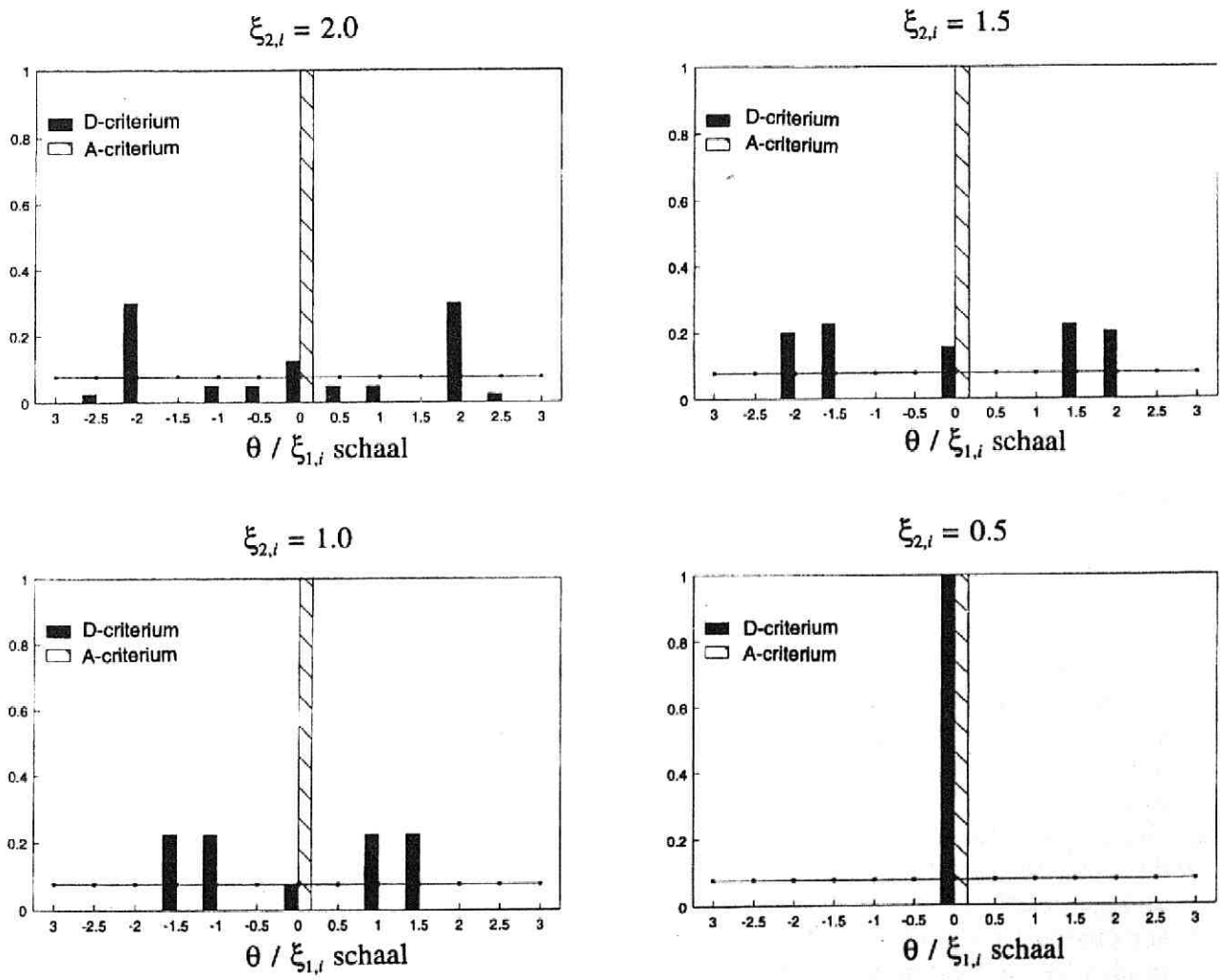


Fig. 3. Optimale tests voor een uniforme vaardigeheidsverdeling samengesteld met het D- en A-optimaliteitscriterium.

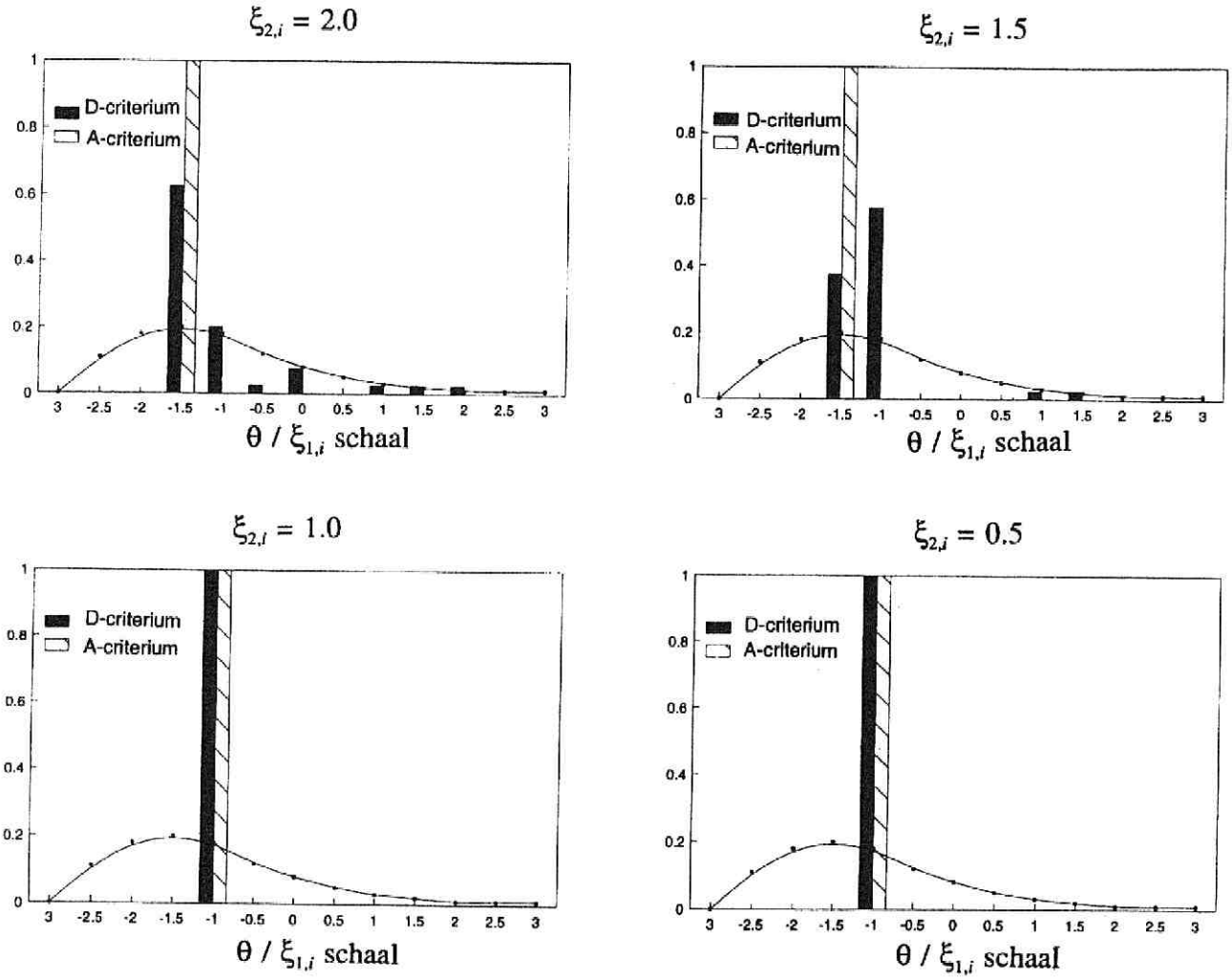


Fig. 4. Optimale tests voor een positief scheve vaardigheidsverdeling samengesteld met het D- en A-optimaliteitscriterium.

In de Figuren 5-7 zijn de resultaten van het ongewogen E-optimaliteitcriterium en het gewogen (maximin) E-optimaliteitscriterium weergegeven. Hieruit kunnen de volgende conclusies getrokken worden:

1. Voor het ongewogen E-optimaliteitscriterium geldt dat de optimale samenstelling van een test zich wijzigt naarmate de discriminatieparameters hogere waarden aannemen. Voor lage discriminatieparameters bestaat een optimaal samengestelde test uit items die bimodaal of trimodaal verdeeld zijn over de moeilijkheidsschaal en voor hogere dicriminatieparameters wordt de optimale verdeling van de items meer uniform van vorm. Het feit dat het ongewogen criterium onafhankelijk is van de gewichten W komt tot uiting in de figuren doordat de optimale samenstelling van een test onafhankelijk is van de vorm van de vaardigheidsverdeling.
2. Het maximincriterium heeft voor hoge discriminatieparameters de neiging om de vorm van de vaardigheidsverdeling te volgen. Voor een normale en een scheve vaardigheidsverdeling en lage discriminatieparameters, namelijk 0,5, 1,0 en 1,5, zal voor dit criterium de optimale samenstelling van een test bestaan uit items met een moeilijkheidsparameter die ligt op de modus van de vaardigheidsverdeling.
3. Voor een uniforme vaardigheidsverdeling zijn de gewichten alle even groot en vertonen het ongewogen en het gewogen E-criterium dezelfde optimale testsamenstelling.

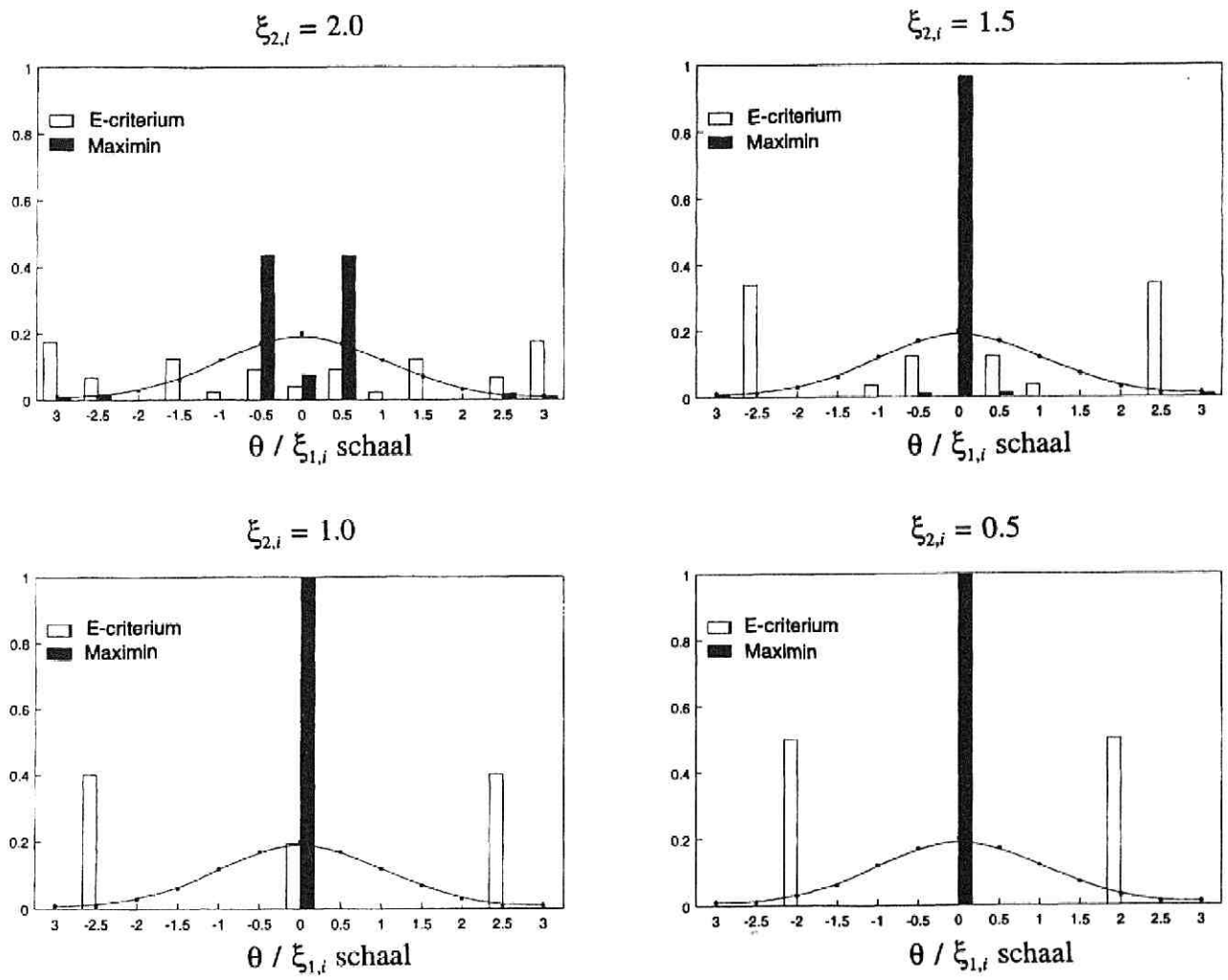


Fig. 5. Optimale tests voor een normale vaardigeheidsverdeling samengesteld met het E- en Maximin optimaliteitscriterium.

DISCUSSIE

In de vorige paragraaf werden de conclusies gepresenteerd uit een aantal analyses met optimaliteitscriteria voor de samenstelling van psychologische tests en studietoetsen. Benadrukt moet worden dat deze conclusies theoretisch van aard zijn en mogelijk waren door de aanname van een ideale itembank zonder beperking in de parameterwaarden van de items. Ook werd in alle analyses verondersteld dat de geselecteerde items geen verschillen in de waarden van hun discriminatieparameters vertoonden. Niettemin hebben de conclusies praktische betekenis. Zo is het bijvoorbeeld een opvallend resultaat dat *geen* van de grafieken een verdeling van items oplevert die vaak op intuïtieve gronden in IRT analyses wordt gevolgd en waarbij wordt gestreefd naar items die over het gehele interval op gelijke afstand van elkaar liggen. Verder blijken de optimale itemverdelingen systematisch af te hangen van de discriminatiewaarden en van de vaardigheidsverdeling in de beoogde populatie van personen. In het algemeen kan gesteld worden dat wanneer de items een lage discriminatiewaarde hebben, een optimale test zal bestaan uit items die een moeilijkheid hebben welke ligt rond de modus van de vaardigheidsverdeling. Voor items met een hoge discriminatieparameter zal de optimale samenstelling meer afhangen van de vorm van de vaardigheidsverdeling. Hier zal een test meestal een optimaal resultaat benaderen wanneer de items een verdeling van de moeilijkheidsparameter hebben, die op de verdeling van de vaardigheden lijkt.

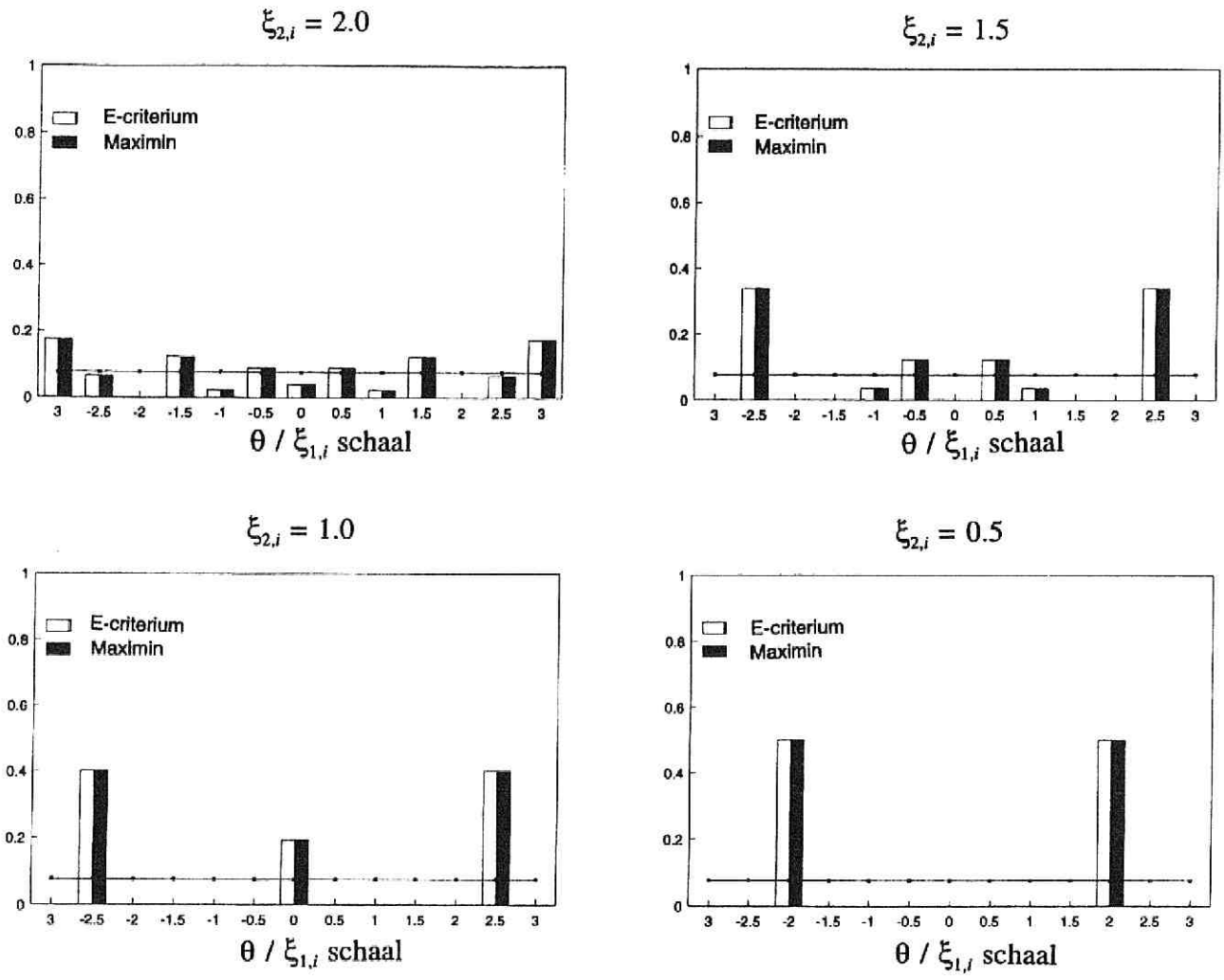


Fig. 6. Optimale tests voor een uniforme vaardigheidsverdeling samengesteld met het E- en Maximin optimaliteitscriterium.

Bij de toepassing van de resultaten op realistische itembanken moet bedacht worden dat bij itemselectie de aanwezigheid van schattingsfouten in de itemparameters tot suboptimale resultaten aanleiding kan geven. Hambleton, Jones en Rogers (1993) hebben laten zien dat door het gebruik van schattingen de testinformatiefunctie systematisch overschat kan worden. Dit verschijnsel kan worden verklaard door kanskapitalisatie. Of dit verschijnsel optreedt en hoe sterk het is, hangt evenwel niet alleen van de grootte van de schattingsfouten af, maar ook van het gehanteerde optimaliteitscriterium en van de eventuele aanvullende randvoorwaarden die aan het criterium worden opgelegd.

LITERATUUR

Atkinson, A.C. (1982). Developments in the design of experiments. *International Statistical Review, 50,* 161-177.

Berger, M.P.F. (1992a). Generation of optimal designs for nonlinear models when the design points are incidental parameters. In Y. Dodge and J. Whittaker (eds.). *Computational Statistics,* Volume 2 (pp. 200-208), New York: Springer Verlag.

Berger, M.P.F. (1992b). Sequential sampling designs for the two-parameters item response theory model. *Psychometrika, 57,* 521-538.

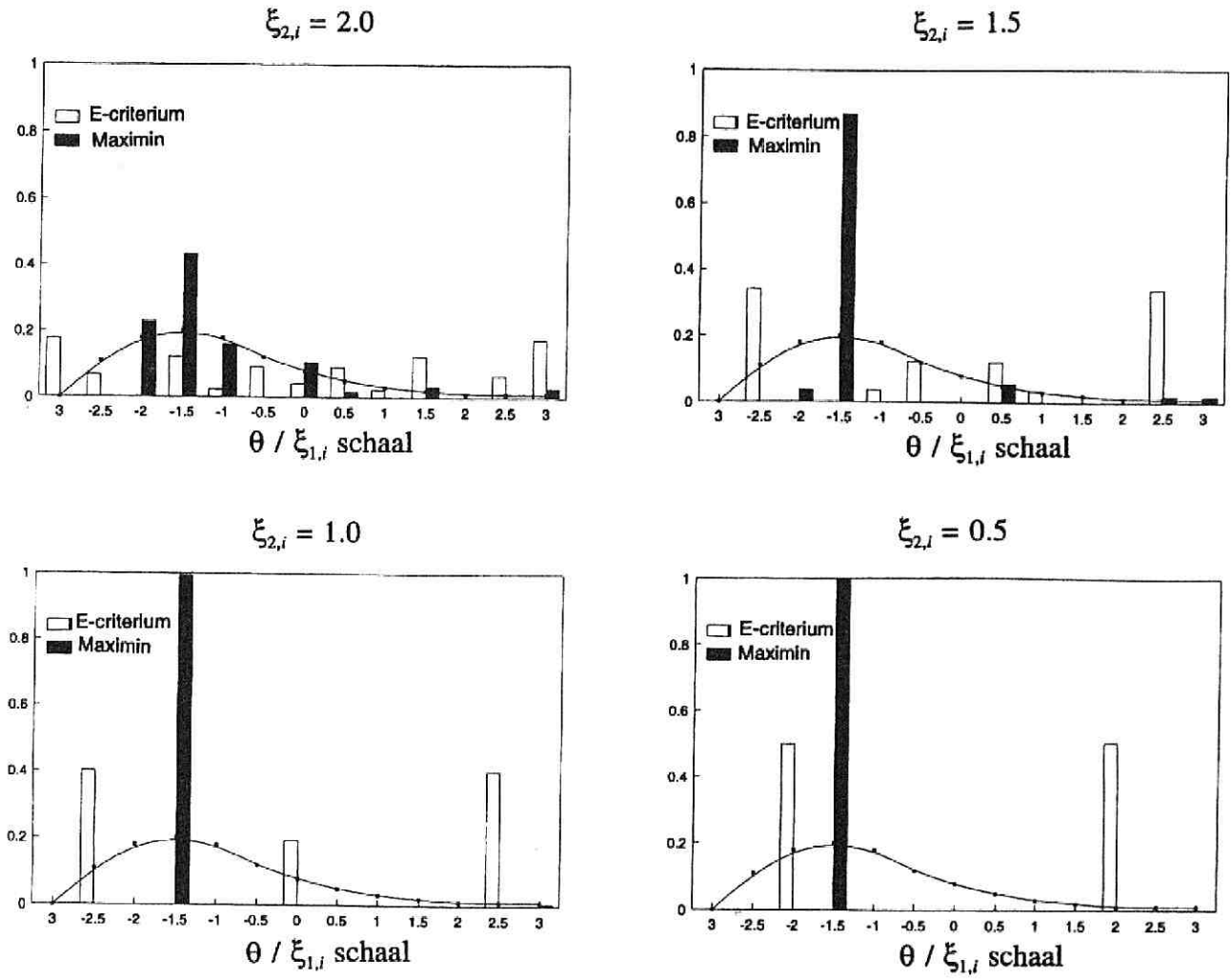


Fig. 7. Optimale tests voor een positief scheve vaardigeheidsverdeling samengesteld met het E- en Maximin optimaliteitscriterium.

Berger, M.P.F. (1994). D-optimal sequential sampling designs for item response theory models. *Journal of Educational Statistics, 19*, 43-56.

Berger, M.P.F. (in press). A general approach to algorithmic design of fixed-form tests, adaptive tests and testlets. *Applied Psychological Measurement.*

Berger, M.P.F. & van der Linden, W.J. (1992). Optimality of sampling designs in item response theory models. In M. Wilson (ed.). *Objective measurement: Theory into practice* (pp. 274-288). Norwood, NJ: Ablex Publishing.

Hambleton, R.K., Jones, R.W., & Rogers, H.J. (1993). Influence of item parameter errors in test development. *Journal of Educational Measurement, 30*, 143-156.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory*. Boston: Kluwer-Nijhoff.

Kalvenlage, E. *PCPROG* [computerprogramma]. Amsterdam: Quantitative Management Software.

Khan, M.K. & Yazdi, A.A. (1988). On D-optimal designs. *Journal of Statistical Planning and Inference, 18*, 83-91.

Kendall, M.G. & Stuart, A. (1973). *The advanced theory of statistics* (vol. 2). New York: Hafner.

Van der Linden, W.J. (1994). Optimal design in item response theory: Applications to test assembly and item calibration. In G.H.Fischer & D. Laming (eds.). *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 303-318). New York: Springer Verlag.

Van der Linden, W.J. & Boekkooi-Timminga, E. (1989). A maximin model for test design with practical constraints. *Psychometrika, 53*, 237-247.

Lord, F.M. (1962). Estimating norms by item-sampling. *Educational and Psychological Measurement, 22*, 259-267.

Pandey, T.N. & Carlson, D. (1976). Assessing payoffs in the estimation of the mean using multiple matrix sampling designs. In D.N.M. de Gruijter & L.J. van der Kamp (eds.). *Advances in psychological and educational measurement*. London: Wiley.

Silvey, S.D. (1980). *Optimal design*. London: Chapman & Hall.

Steinberg, D.M. & Hunter, W.G. (1984). Experimental design: Review and comment. *Technometrics, 26*, 71-130.

Stocking, M. (1990).Specifying optimum examinees for item parameter estimation in item response theory. *Psychometrika, 55*, 461-475.

Theunissen, T.J.J.M., Sanders, P.F. & Verschoor, A.J. (1993). Het samenstellen van toetsen. In T.J.H.M Eggen & P.F. Sanders (eds.). *Psychometrie in de praktijk* (pp. 409-442). Arnhem: Cito.

Veerkamp, W.J.J. & Berger, M.P.F. (1993). Some new item selection criteria for adaptive testing. Research Rapport, Enschede: University of Twente.

Verhelst, N.D. & Eggen, Th.J.H.M. (1989). *Psychometrische en statistische aspecten van peilingsonderzoek*. Arnhem: CITO, PPON-rapport nr. 4.

Manuscript ontvangen 16-5-1994
Definitieve versie ontvangen 23-8-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, pp. 93-95

Notities en Commentaren

EEN "BOEKBESPREKING" ALS BOEMERANG: BLOK OF HET WETENSCHAPPELIJK FORUM

T. Mooij
KU/ITS Nijmegen

1. INLEIDING

In zijn bespreking van het onderzoeksrapport: "Kenmerken en effecten van methoden begrijpend/studerend lezen" gaat Blok (zie Tijdschrift voor Onderwijsresearch, 1994, 19, 4, 375-380) eerst in op het waarom van de boekbespreking. Er zou, ten eerste, in het basisonderwijs veel tijd besteed worden aan begrijpend lezen en de effecten van dit leesonderwijs zouden in sterke mate bepalen hoe goed leerlingen in het voortgezet onderwijs in staat zijn om zelfstandig kennis uit studieboeken te verwerven. In de tweede zin die hierop volgt wordt, ten tweede, evenwel gemeld dat er veel te weinig instructie gegeven zou worden. Ten derde zou het besproken onderzoek min of meer model staan voor ander methode-vergelijkend onderzoek dat door SVO geëntameerd is.

Bij het lezen van de bespreking zelf wordt echter duidelijk dat het Blok niet zozeer gaat om het begrijpend/studerend lezen of om de verschillende effecten hiervan op leerlingen. Hij zou dan bijvoorbeeld toe moeten geven dat effecten vanuit het basisonderwijs tot in het voortgezet onderwijs niet zijn aangetoond (zie het eerste punt van het waarom). Hij zou bovendien moeten vermelden dat het onderhavige onderzoek aantoont dat een bepaalde categorie leerlingen juist te veel instructie in begrijpend/studerend lezen krijgt (tweede punt van het waarom). Blijft over het derde punt: het gaat blijkbaar om SVO.

Toch lijkt er sprake van een tweede reden voor de bespreking door Blok. Deze betreft het bezig zijn met begrijpend/studerend lezen in institutioneel verband inclusief de bijbehorende concurrerende verhoudingen tussen (personen in) onderzoeksinstituten die hun "markt" moeten uitbreiden of verdedigen. Dát Mooij in dit verband extra aandacht van Blok zou krijgen, werd door Blok in april 1994 al aangekondigd: "Laat eerst de wetenschappers maar eens beoordelen, hoe netjes Mooij zijn werk gedaan heeft" (zie Didaktief, april 1994, p. 14). Wetenschapper Blok voegt nu zelf de daad bij het woord. De positie van SCO (het instituut van Blok) op de markt van het begrijpend/studerend lezen zou, gezien de "boekbespreking", gediend zijn met het ter discussie stellen van het functioneren van SVO door middel van kritiek op Mooij (het ITS). Twee vliegen in één klap.

Opvallend echter is dat de mening van Blok tot nu toe in geen enkel opzicht overeenkomt met die van andere leden van het wetenschappelijk forum. Integendeel, in een eerder stadium is zijn mening zelfs uitdrukkelijk afgewezen door het forum zonder dat Blok dit in de "boekbespreking" noemt. Om voor iedere geïnteresseerde een juiste beoordeling van het genoemde onderzoek mogelijk te maken, wil ik hierbij beknopt reageren op de "boekbespreking" van Blok.

2. VOORGESCHIEDENIS

Eind 1990 vraagt SVO aan elk van de onderzoeksinstituten SCO, RION en ITS een onderzoeksvoorstel in te dienen in het kader van de SVO-aanbesteding. Het onderwerp is de evaluatie van begrijpend/studerend leesmethoden. Er dienen vier door SVO gespecificeerde onderzoeksvragen beantwoord te worden. De gebruikelijke procedure bij deze aanbesteding is dat de instituten

Adres: Postbus 9048, 6500 KJ Nijmegen.

in onderlinge concurrentie een voorstel kunnen indienen. Elk voorstel wordt door een wetenschappelijke commissie beoordeeld en uiteindelijk zal één instituut de opdracht tot het uitvoeren van het onderzoek kunnen verwerven. Wijziging van de door SVO gestelde randvoorwaarden (vraagstellingen, onderzoeksopzet, beschikbare financiën, tijdsduur) is in dit stadium niet meer mogelijk.

Begin januari 1991 ben ik (binnen het ITS) telefonisch benaderd door Blok (van SCO). Zijn mededeling was dat zijns inziens de vraagstellingen van het gevraagde evaluatie-onderzoek niet deugden. Zijn vraag was of er tussen instituten niet gezamenlijk opgetreden kon worden om één en ander te veranderen. Mijn reactie hield in dat ik ook van mening was dat de vraagstellingen beter zouden kunnen, maar dat de ervaring leerde dat er bij de uitwerking van het onderzoeksvoorstel niets meer aan te veranderen zou zijn. Toch stelde ik voor om binnen RION, SCO en ITS na te gaan of de directies van de instituten bereid waren tot gezamenlijke actie naar de SVO-directie toe. De SVO-directie zou de vraagstellingen kunnen veranderen. Blok was het hiermee eens en ging dit na binnen SCO. Ik zelf ben dit nagegaan binnen RION én ITS. De uitslag van dit onderzoekje past in het beeld van "concurrentie": de directies wilden of konden niet tot één gezamenlijke strategie komen.

Uiteindelijk diende elk instituut een eigen onderzoeksvoorstel in. Na de indieningstermijn stuurden de onderzoekers elkaar hun voorstel toe: een uiting van solidariteit. Het bleek dat Blok een onderzoeksvoorstel had dat afweek van de door SVO gestelde randvoorwaarden.

De in de procedure gebruikelijke, van SVO onafhankelijke, wetenschappelijke commissie ter beoordeling van onderzoeksvoorstellen karakteriseerde het ITS-voorstel als het beste voorstel. SVO nam dit oordeel over en gunde het onderzoek aan het ITS. Met andere woorden: het wetenschappelijk forum gaf de voorkeur aan het onderzoeksvoorstel van Mooij boven dat van Blok. Daarop verleende SVO aan het ITS de faciliteiten om het onderzoek volgens de randvoorwaarden uit te voeren.

3. HET EINDRAPPORT

In de bespreking van Blok wordt de beoordeling van het eindrapport door SVO omschreven met: "(...) dat de SVO in haar rol als opdrachtgever heeft verklaard, dat het onderhavige rapport aan haar minimum kwaliteitseisen voldoet.".

SVO evenwel kan geacht worden een vergelijking te kunnen maken tussen de beschikbaar gestelde randvoorwaarden en het geleverde produkt. In de SVO-beoordeling van het eindrapport staat onder andere wat betreft:

. technisch ontwerp: "Een en ander wordt helder uiteengezet en geïllustreerd (...)";
. beschrijving instrumenten: "De uitgebreide beschrijvingen zijn helder." en: "kwaliteit ok";
. analyse: "adequaat";
. weergave resultaten: "duidelijk";
. algemene presentatie en leesbaarheid: "helder en informatief".

Blok kent in zijn "boekbespreking" geen inperking via SVO-randvoorwaarden: hij eist informatie ook wanneer die, gezien de grote lijn in de onderzoeksresultaten, niet zinvol is. Toch zijn er, ook buiten SVO-randvoorwaarden om, beperkingen die zelfs voor Blok gelden. Bijvoorbeeld: hoewel hij dit nu als negatief punt van Mooij opvoert, zou ook Blok de methode "Wie dit leest" voor wat betreft de groepen 7-8 in het schooljaar 1991-1992 niet hebben kunnen gebruiken. Deze methode was toen nog niet gereed voor deze groepen. Voorbeeld twee: Blok wil enerzijds een verdergaande inhoudelijke analyse van methoden. Op zich juist, maar is zo'n analyse het werk van de onderzoeker in het kader van een effectenonderzoek? Anderzijds wil Blok tegelijkertijd ook grootschaliger en langduriger effectenonderzoek bij leerlingen dan nu uitgevoerd is. Blok heeft blijkbaar de beschikking over randvoorwaarden die dit alles mogelijk maken. De vraag is dan: waar blijft het betreffende onderzoek? Blok lijkt in dit verband ook niet echt te beseffen dat er in een lopend onderzoek inclusief rapportage voortdurend keuzen gemaakt

moeten worden. Bijvoorbeeld: de informatie van Laarhoven en van Van Rossum heb ik, in tegenstelling tot wat Blok beweert, wel gebruikt bij het opstellen van de variabelenlijst ter inhoudelijke scoring van methoden. Dit had hij kunnen nagaan in het rapport (p. 45-47). Omdat scoring vanwege de in het rapport vermelde redenen niet mogelijk bleek, is de methodenanalyse uiteindelijk vervallen. Ook zonder inhoudelijke methodenanalyse zijn methodeneffecten op leerlingen vast te stellen. Blok echter, neem ik aan, zou in die situatie gestopt zijn met het gehele onderzoek.

4. HET WETENSCHAPPELIJK FORUM

Ik bied altijd zelf mijn werk ter wetenschappelijke beoordeling aan. Leden van het wetenschappelijk forum kan ik wijzen op het artikel "Methode- en groepseffecten op leerprocessen in begrijpend lezen" in het themanummer: "Klassesamenstelling: processen en effecten" van het Pedagogisch Tijdschrift (1994, 19(2), p. 143-153). Elk artikel in dit nummer van maart/april 1994 is beoordeeld vanwege de speciale themanummer-redactie. Tevens worden de gepresenteerde artikelen door een ander lid van het forum nabesproken. In tegenstelling tot "wetenschapper Blok" vindt het forum het bedoelde onderzoek blijkbaar wel wetenschappelijk verantwoord en van belang.

5. CONCLUSIES

Als het Blok in zijn "boekbespreking" gaat om SVO, ofwel om de structurering of ook financiering van onderwijsonderzoek, dan is zijn bespreking een zwaktebod. Voor gefundeerde kritiek op het functioneren van SVO is selectie van slechts één onderzoek en één persoon en één onderzoeksinstituut onvoldoende. Deze handelwijze wekt in sterke mate de indruk dat Blok de SCO wat betreft het begrijpend/studerend lezen wil bevoordelen. Dit wordt benadrukt doordat Blok de voorgeschiedenis van dit project, en vooral zijn eigen betrokkenheid daarin, niet noemt in de "boekbespreking". Mijns inziens dienen structurele aspecten betreffende het onderwijsonderzoek op de juiste wijzen en op de juiste niveaus behandeld te worden. Bovendien heeft SVO dan de gelegenheid om de eigen zienswijzen en standpunten naar voren te brengen, zoals dit behoort te kunnen.

Het lijkt Blok echter (vooral) te gaan om het werk van Mooij op het terrein van het begrijpend/studerend lezen, gezien zijn uitspraak: "Laat eerst de wetenschappers maar eens beoordelen, hoe netjes Mooij zijn werk gedaan heeft". Wat dit betreft kan er op grond van het bovenstaande een heldere conclusie getrokken worden. Enerzijds ligt er het oordeel van "wetenschapper Blok" in de vorenstaande "boekbespreking". Zijn oordeel is negatief. Anderzijds is mijn werk in eerste instantie beoordeeld door de van SVO onafhankelijke, wetenschappelijke commissie in het kader van de SVO-aanbesteding. Mijn onderzoeksvoorstel werd beoordeeld als kwalitatief beter dan het voorstel van Blok. Ten tweede is het eindrapport van het onderzoek in de SVO-beoordeling omschreven als helder en informatief. Ten derde heeft het wetenschappelijk forum via het "Pedagogisch Tijdschrift" de kern van mijn onderzoek in de vorm van de multi-niveau analyses en -conclusies gepubliceerd. *Tot nu toe staat Blok dus strikt alleen in zijn negatieve oordeel en is het wetenschappelijk forum unaniem positief over mijn werk.* Het door Blok te hulp roepen van "de wetenschappers" werkt dus als een boemerang.

Ontvangen 11-11-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 1, p. 96

Notities en Commentaren

NEUZEN TELLEN?

H. Blok
SCO-Kohnstamm Instituut voor Onderzoek van Opvoeding en Onderwijs van de Universiteit van Amsterdam

Ik betreur het oprecht dat mijn boekbespreking Mooij tot het schrijven van zijn gepikeerde reactie heeft gebracht. In mijn bespreking heb ik mijn best gedaan de kritiek op zijn onderzoek te herleiden tot de randvoorwaarden – gesteld door de SVO – waaronder hij zijn onderzoek heeft moeten verrichten. Die randvoorwaarden staan soms wetenschappelijk en maatschappelijk relevant onderzoek danig in de weg, zo heb ik trachten duidelijk te maken. Met andere woorden, niet in de eerste plaats Mooij maar de SVO kan zich aangevallen voelen. Wat ik met mijn bespreking tot stand hoopte te brengen – die hoop is nog niet vervlogen – is een discussie onder vakgenoten over die randvoorwaarden. Ik ben ervan overtuigd dat Mooij aan deze discussie een zinvolle bijdrage zou kunnen leveren.

Helaas voert Mooij in zijn weerwoord argumenten aan die bij een andere discussievraag horen. Zo zou mijn bespreking ingegeven zijn door teleurstelling over het feit dat de onderzoeksopdracht niet aan mij maar aan Mooij is verstrekt. Als dat al waar zou zijn, lijkt het mij niet relevant voor een zakelijke discussie. Maar Mooij vergist zich in mijn betrokkenheid. Ik heb destijds geen concurrerend onderzoeksvoorstel geschreven. In het slot van zijn reactie doet Mooij het voorkomen alsof het wetenschappelijk forum slechts neuzen hoeft te tellen. Dat moge op de korte termijn soms zo lijken, maar op de lange termijn zijn niet zo zeer de neuzen als wel de argumenten doorslaggevend.

Ontvangen 5-12-1994

Adres: Grote Bickersstraat 72, 1013 KS Amsterdam.

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 2, pp. 97-114

Themanummer: organisatie en management van de meer autonome school

Inleiding: Nuances in autonomie

J. Scheerens
Universiteit Twente, Enschede

ABSTRACT

School autonomy and the policy of devolution of authority to the school level is a phenomenon with many features. International comparisons show that, at least in the primary and secondary sector, school autonomy in the Netherlands is already quite considerable. In these sectors there appears to be relatively little room for further enlargement of school autonomy, particularly when the position is taken that core educational objectives ought to be centrally established (and possibly assessed). At the middle level of vocational education other partners, apart from the state (e.g. branch organisations of private industries) participate in the establishment of educational objectives, whereas institutes at the level of higher and university education are at liberty in choosing educational objectives. The autonomy of schools in the Netherlands in the area of educational processes has traditionally been quite large. At all levels more discretion in the areas of finance, personnel management and organization and management is planned and partially being implemented.
A major research-question that can be geared to this educational policy is the question whether more autonomy in the areas of general management, budgetary control and personel management enhances school-effectiveness and a more responsive and innovative behaviour of schools.

INLEIDING

Sinds het eind van de jaren tachtig is er in het onderwijsbeleid sprake van een hernieuwd enthousiasme voor het vergroten van de autonomie van scholen. Dit blijkt ondermeer uit een serie discussie- en beleidsnota's, met als voorbeelden "Hoger Onderwijs: Autonomie en Kwaliteit", "De school op weg naar 2000" en de "Gezamenlijke richtinggevende uitspraken" van het zogenoemde Scheveningse Beraad Bestuurlijke Vernieuwingen. "Hernieuwd", omdat de inrichting van het Nederlandse onderwijsbestel traditioneel bepaald is door principes als soevereiniteit in eigen kring en het subsidiariteitsbeginsel en de ideeën over een relatief autonome school ook al eerder bepleit werden. Nu de accentuering van autonomie van scholen zich in de beleidsintenties ook verder uitstrekt dan het pedagogisch-didactische domein, maar zich ook juist richt op bekostigingsvoorwaarden, bestuurlijk-organisatorische aspecten en arbeidsverhoudingen is de vraag aan de orde naar de implementatie ervan in de onderwijspraktijk. Op deze vraag wordt in enkele bijdragen aan dit themanummer voor bepaalde aspecten van autonomievergroting een empirisch antwoord gegeven. Overigens valt uit persberichten, waarin nog steeds sprake is van Haagse "regelwoede", op te maken dat, naar de perceptie van betrokken geledingen uit het onderwijsveld, de deregulering nog niet echt gestalte heeft gekregen.

Uit recente bijdragen over het onderwerp in kwestie (b.v. Leune, 1994, Van Wieringen, 1994) wordt duidelijk dat de "autonomisering" van de Nederlandse onderwijsinstellingen selectief van aard is en het best voor iedere onderwijssector afzonderlijk bekeken kan worden, gezien aanmerkelijke sectorspecifieke verschillen. Zelfs moet de vraag gesteld worden naar het retorisch gehalte van de vele deliberaties over autonomievergroting en deregulering, waar onder-

* Adres: Faculteit Toegepaste Onderwijskunde, Universiteit Twente, Postbus 217, 7500 AE Enschede.

meer geconstateerd wordt dat op het vlak van de kwaliteitsbewaking de regulering toeneemt. Als voorloper op enkele artikelen waarin verslag wordt gedaan van empirisch onderzoek naar de mogelijkheden en beperktheden die in scholen optreden op het vlak van organisatie en management bij toenemende autonomie is het daarom dienstig om enkele nuances van de veronderstelde autonomievergroting te belichten.

In de eerste plaats is begripsverheldering wenselijk; wat wordt er precies bedoeld met deregulering en autonomievergroting, in hoeverre zijn deze begrippen verwant met definities van decentralisatie, delegatie en mandatering? Daarbij moet, in navolging van Leune (ibid) onderscheid gemaakt worden in verschillende soorten autonomie.

In de tweede plaats is het interessant om de verdeling van beslissingsbevoegdheden over bestuurlijke niveaus in het Nederlandse onderwijssysteem te vergelijken met andere landen.

Ook dit kan aanleiding geven tot nuancering van de beeldvorming: hoe "geautonomiseerd" zijn Nederlandse scholen eigenlijk bezien in internationaal perspectief?

De meest intrigerende vraag blijft die naar de relatie tussen autonomie en kwaliteit van het onderwijs. Deze vraag is eerder ondermeer behandeld door Scheerens, 1993 en door Creemers, 1994. Kernpunten daarbij zijn de definitie van kwaliteit die wordt gehanteerd, de eventuele kwaliteitsbevorderende impulsen van veranderde organisatie-condities bij autonomere scholen en de theorievorming daaromtrent.

Aan deze aspecten van de beleidsmatige context van meer autonome scholen zal kort aandacht worden besteed, met als doel de schoolgebonden ontwikkelingen die in de overige bijdragen centraal staan te kunnen situeren.

BEGRIPSBEPALING

Bij deregulering is er in het algemeen sprake van minder en minder gedetailleerde regelgeving vanuit een hoger bestuursniveau naar een lager bestuursniveau. Hierdoor neemt de speelruimte of de autonomie van het lagere bestuursniveau ten opzichte van het hogere bestuursniveau toe. Toch is een meer precieze typering gewenst. Van Wieringen (1994) noemt twee verschijningsvormen, die meer inzicht bieden in de wijze waarop deregulering werkt. De eerste is dat deregulering, gedacht binnen een hiërarchie van besturingsniveaus, het karakter heeft van sturing die aangrijpt op een hoger systeemniveau. Een voorbeeld is dat regelgeving in het onderwijs zich meer concentreert op het niveau van het bevoegd gezag dan dat van de individuele docent. Het tweede type verschijningsvorm van deregulering is besturing via omgevingsactivering. In deze situatie lopen de stuurmaatregelen vanuit het hoogste besturingsniveau niet rechtstreeks naar de te besturen organen op lager niveau, maar via de band van de omgeving. Voorbeelden in het onderwijs zijn het stimuleren van actief consumentengedrag van ouders in het primaire onderwijs en convenanten met de sociale partners in het secundair beroepsonderwijs (Van Wieringen, ibid).

In de praktijk komt deregulering van het onderwijssysteem neer op een toespitsing van centrale bemoeienis op de grondwettelijk vastgelegde kerntaken van de overheid en op het scheppen van condities, waarbinnen scholen zoveel mogelijk eigen verantwoordelijkheid kunnen dragen (Min. van O&W, 1994).

Decentralisatie, preciezer gezegd verticale decentralisatie, in de zin van het verleggen van formele beslissingsbevoegdheden van hogere naar lagere bestuursniveaus, ligt in het verlengde van deregulering. Ook al behoeft er bij deregulering strikt genomen geen formele overdracht van bevoegdheden plaats te vinden, en zou men de mate van regulering als een besturings"stijl" kunnen opvatten, toch betekent deregulering altijd meer speelruimte (en verantwoordelijkheid) voor een lager besturingsniveau. "Mandateren" gaat in deze minder ver dan verticaal decentraliseren, omdat er bij mandateren alleen sprake is van een overdracht aan een lagere instantie, terwijl de hogere instantie wel verantwoordelijk blijft (De Leeuw, 1982, p. 244). Een voorbeeld uit het onderwijs is de mandatering van bevoegdheden vanuit het bevoegd gezag naar de directie

of schoolleiding (vgl. de Gezamenlijke richtinggevende uitspraken van het Schevenings Beraad Bestuurlijke Vernieuwingen, 1993, p. 35). Ook horizontale decentralisatie is aan de orde bij de huidige tendens tot deregulering in het onderwijs. Bij horizontale decentralisatie gaat het niet alleen om formele beslissingsbevoegdheid, maar tevens om de algemene autonomie en zelfstandigheid van onderdelen van een organisatie, c.q. het onderwijsbestel (De Leeuw, 1982, p.241).

Mintzberg (1979, p. 185) omschrijft horizontale decentralisatie als de mate waarin "nonmanagers" invloed hebben op besluitvorming. Relatief autonome studierichtingen of secties binnen grote scholen zijn als voorbeeld te noemen. Een derde vorm van decentralisatie, ook wel aangeduid als deconcentratie oftewel een grotere geografische spreiding van eenheden, neemt in het onderwijs, door de koppeling van autonomisering aan schaalvergroting eerder af dan toe.

Met het enigszins begripsmatig op orde stellen van het geheel aan onderling verwante verschijnselen dat schuilgaat onder de noemers van deregulering en autonomievergroting van scholen is nog niet aangegeven om welke aspecten van het functioneren van het onderwijsbestel het hierbij gaat. Een recentelijk door Leune (1994) gemaakte classificatie van schoolautonomie biedt hierbij uitkomst.

SELECTIEVE AUTONOMIEVERGROTING

Leune onderscheidt de volgende soorten van school-autonomie:

- budgettaire autonomie, die betrekking heeft op zaken als huisvesting, de besteding van verkregen financiële middelen en de administratie;
- bestuurlijke en organisatorische autonomie, betreffende de inrichting van bestuur en organisatie van de school;
- normatieve autonomie, c.q. de vrijheid om de normatieve richting van de school te bepalen;
- arbeidsvoorwaardelijke autonomie, die zowel betrekking heeft op personeelsbeleid op rechtspositioneel gebied als op het personeelsbeleid inzake de professionele bekwaamheid van leraren en schoolleiders;
- onderwijskundige autonomie, met daarbij het belangrijke onderscheid tussen procesmatige en produktmatige autonomie, waarbij de produktmatige autonomie betrekking heeft op de na te streven einddoelen, eindtermen en examencisen.

Gebruik makend van deze taxonomie beziet Leune het vigerende beleid van de rijksoverheid op het terrein van de bestuurlijke verhoudingen op z'n merites, waarbij hij zich beperkt tot het primair en voortgezet onderwijs. Hij concludeert daarbij dat er inderdaad sprake is van autonomie-vergroting op budgettair, rechtspositioneel en procesmatig onderwijskundig terrein, terwijl de vergroting van autonomie op arbeidsvoorwaardelijk-professioneel terrein nog volop in discussie is. Verder stelt hij dat de bestaande autonomie op het bestuurlijk-organisatorisch vlak niet betekenisvol verandert, terwijl ook de bestaande normatieve autonomie niet als beginsel ter discussie staat. De produktmatige onderwijskundige autonomie is door de invoering van kerndoelen eerder kleiner geworden.

Hierbij zou de kanttekening geplaatst kunnen worden dat de procesmatige onderwijskundige autonomie altijd al groot is geweest en dat deze door de invloed van kerndoelen en eindtermen op het vlak van de selectie van het onderwijsaanbod juist weer enigszins ingeperkt zou kunnen worden. Verder blijft met name de invulling van het schoolleiderschap (als onderdeel van de bestuurlijk-organisatorische verhoudingen) niet onberoerd, bijvoorbeeld door een toegenomen mandaat vanuit het bevoegd gezag, als neveneffect van een grotere budgettaire autonomie. In hoeverre dit, in het basisonderwijs, nu reeds tot veranderingen heeft geleid komt aan de orde in de bijdrage van Stoel.

In de "hogere" onderwijssectoren wijkt de realisering van de verschillende door Leune onderscheiden soorten autonomie op onderdelen behoorlijk af. Dit zal worden toegelicht aan de hand van drie ideaaltypische modellen van de inrichting van onderwijsstelsels die behandeld worden in een recent rapport van de WRR over de aansluiting van onderwijs en arbeid (Derck-

sen et al. 1993). De drie onderscheiden modellen zijn respectievelijk het *centraal planningsmodel van onderwijs*, het *Angelsaksische model van de zelfstandige school* en het *continentale model van de zelfstandige school.*

Het centrale planningsmodel van onderwijs wordt gekenmerkt door het van overheidswege vastleggen van regels en beleid op onderwijsinhoudelijk en financieel terrein. Het aanbod is vastgelegd in de vorm van door de minister goedgekeurde leerplannen, terwijl landelijke examens uniform de te bereiken kwalificaties vastleggen. Ook de overheidsfinanciering van het onderwijs verloopt via vaste regels, bijvoorbeeld door middel van een declaratiestelsel. Het management heeft sterk het karakter van ambtelijke uitvoering van centraal gestelde regels. De arbeidsvoorwaarden van het personeel zijn gelijk aan die van andere ambtenaren.

Het Angelsaksische model van de zelfstandige school komt overeen met de wijze waarop het universitaire onderwijs in de Verenigde Staten is ingericht. De leerlingen of studenten zijn zelf verantwoordelijk voor de keuze van hun opleiding. De onderwijsinstellingen bepalen het onderwijsaanbod en concurreren met elkaar om leerlingen, door een aantrekkelijk aanbod te doen, kwalitatief goed onderwijs te bieden of te schermen met arbeidsmarktvooruitzichten. De financiering van het onderwijs komt tot stand door bijdragen van de leerlingen. Een voucher systeem is hiervan een voorbeeld. Instellingen kunnen door middel van eigen beurzen bepaalde categorieën leerlingen aantrekken, en hebben tevens de vrijheid om selectienormen te hanteren. In dit model is er geen of hoogstens een gedeeltelijke overheidsfinanciering. Wel is denkbaar dat behalve de leerlingen ook het bedrijfsleven als (betalend) consument optreedt. Het instellen van

Tabel 1. Drie ideaaltypische modellen van de inrichting van onderwijsstelsels (naar Dercksen et al. 1993, p. 45-49).

	onderwijsinhoud	financiering	management
Centrale planningsmodel onderwijs	. centrale vaststelling opleidingenaanbod . landelijke eindtermen (c.q. leerplan en centrale examens)	. overheidsbekostiging en/of individuele bekostiging	. ambtelijke uitvoering van centraal gestelde regels . duur van opleidingen is gegeven . arbeidsvoorwaarden conform die van (andere) ambtenaren
Angelsaksische model zelfstandige school	. scholen stellen eigen opleidingenaanbod vast . scholen bepalen eindtermen zelf	. (zeer) hoge eigen bijdragen (bijv. via persoonlijke leningen) . prestatiebeurzen . partnerships met bedrijven	. ondernemend op alle fronten . school bepaalt duur van opleidingen . school bepaalt (onderhandelt) arbeidsvoorwaarden zelf
Continentale model zelfstandige school	. centrale vaststelling opleidingenaanbod . landelijke eindtermen	. overheidsbekostiging via outputfinanciering . studiefinanciering . (eventueel) bescheiden eigen bijdragen	. selecteert aan te bieden opleidingen . ondernemend m.b.t. uitvoering onderwijs . school bepaalt opleidingsduur . arbeidsvoorwaarden volgens CAO van sector onderwijs, c.q. ondernemings-CAO

scholen of opleidingen is in dit ideaaltypische model aan weinig regels gebonden. Voor het management geldt dat het in principe niet verschilt van het leiden van een onderneming. De arbeidsvoorwaarden worden door de instelling bepaald of geregeld via een eigen CAO. Tenslotte ligt het in de rede dat bij de toepassing van dit model in het beroepsonderwijs de neiging zal bestaan zich sterk te oriënteren op het lokale bedrijfsleven en te komen tot een-op-een relaties van scholen en bedrijven, dan wel van scholen en branche-organisaties.

Het derde ideaaltypische model, het continentale model van de zelfstandige school, wordt vooral bepaald door controle op de output, outputfinanciering en liberalisatie van de onderwijskundige proceskant. De vrijheid van de scholen om het onderwijsaanbod te bepalen is wel aanwezig, maar toch enigszins beperkt: scholen kiezen uit een beperkt aantal van hoger hand vastgestelde opleidingen. Overheid en bedrijfsleven zijn beide partner bij de vaststelling van eindtermen en de limitering van het onderwijsaanbod. Essentieel in dit model is verder dat er, door middel van landelijk equivalente examens, een uniforme en onafhankelijke toetsing van de te bereiken kwalificaties (c.q. eindtermen) mogelijk is. In dit model concurreren scholen niet op het vlak van de inhoud van het onderwijs, maar uitsluitend op basis van performance.

In de onderstaande tabel, ontleend aan Dercksen et al., (1993, p. 45-49) worden de hoofdkenmerken van de drie modellen samengevat.

In zuivere vorm komt geen van de drie modellen overeen met de bestaande situatie in Nederland in de diverse onderwijssectoren. De situatie in het basis- en voortgezet onderwijs komt in grote trekken overeen met het centrale planningsmodel. Daarbij zijn dan wel drie kanttekeningen te plaatsen. In de eerste plaats is er geen sprake van een centrale vaststelling van het onderwijsaanbod, in de vorm van nationale curricula. In de tweede plaats zijn de eindtermen in het basisonderwijs en de eerste fase van het voortgezet onderwijs in algemene termen gesteld en (nog) niet tot uniforme eindtoetsen geoperationaliseerd. In de derde plaats biedt de tendens naar lump-sum financiering meer speelruimte voor de scholen, hetgeen tezamen met enkele voorzichtige stappen in de richting van meer mogelijkheden tot eigen personeelsbeleid, appelleert aan een wat minder ambtelijke vorm van management. Opvallend is dat de in verschillende recente beleidsnota's bepleite zelfevaluatie van scholen, in de vorm van bijvoorbeeld openbare jaarverslagen in feite passen in het meer radicale Angelsaksische verzelfstandigingsmodel, hoewel er daarnaast ook meer centralistische toetsing op eisen van deugdelijkheid wordt bepleit (Vgl. de Gezamenlijke richtinggevende uitspraken van het zogenoemde Scheveningse Beraad Bestuurlijke Vernieuwingen).

Het HBO en WO hebben enkele kenmerken gemeen met het Angelsaksische model, met name de mogelijkheid om zelf het opleidingsaanbod en de eindtermen te bepalen. Ingangsselectie vindt plaats bij tweede fase-opleidingen (AIO-netwerken en onderzoekscholen). Ook zijn er in deze sectoren wat meer vrijheidsgraden op arbeidsvoorwaardelijk terrein. Het belangrijkste punt waarop de situatie in Nederland in deze onderwijssectoren afwijkt van het desbetreffende ideaaltypische model is de overheidsfinanciering. Ook de afgedwongen schaalvergroting in het HBO, en trouwens ook in het MBO, is niet te rijmen met de eigen verantwoordelijkheid die scholen, zowel in het Angelsaksische als het continentale model hebben voor respectievelijk het marktconforme aanbod en de output van het onderwijs. Voor zover het al plaats vindt wordt het proces naar grotere autonomie van onderwijsinstellingen in Nederland met een zwaar centralistische ingreep als gedwongen schaalvergroting gestuurd.

De ontwikkelingen in het MBO komen op onderdelen overeen met het continentale model, met name wat betreft de outputgerichtheid: eindtermen die in samenspraak met het bedrijfsleven zijn opgesteld en aanzetten tot rendementfinanciering. Dercksen et al. (1993, p.50) concluderen echter dat de eindtermen bezien tegen de achtergrond van het Angelsaksische en het continentale model vlees noch vis zijn. In het Angelsaksische model zijn ze overbodig, terwijl ze onvoldoende houvast bieden voor toetsing, bezien vanuit het perspectief van het continentale model.

Aan elk van de drie zojuist besproken modellen zitten sterke en zwakke kanten. Zekerheid over een weloverwogen kwalificatiestructuur is het sterke punt van het centrale planningsmo-

del, terwijl flexibiliteit het sterke punt is van het Angelsaksische model. Het continentale model neemt een tussenpositie in. Selectieve autonomisering van het onderwijsbestel kan ook gezien worden als het vinden van een passende gradatie van centrale regulering en zelfstandigheid van instellingen voor elk van de onderwijssectoren. Het eerste criterium (zekerheid over de kwalificatiestructuur) weegt daarbij zwaarder voor het funderend onderwijs dan voor het beroepsonderwijs, terwijl voor het flexibiliteitscriterium het omgekeerde geldt. Aldus beschouwd zouden er voor het funderend onderwijs (basis- en voortgezet onderwijs) slechts lichte modificaties (meer in de zin van partiële deregulering dan verticale decentralisatie) van het centraal planningsmodel wenselijk geacht kunnen worden. Het continentale model zou richtinggevend kunnen zijn voor het MBO en HBO, terwijl er alleen voor het WO te filosoferen zou zijn over verdere overeenstemming met het Angelsaksische model. Daarbij moet dan nog wel de kanttekening worden gemaakt, dat de marktgerichte modellen sterk kapitaliseren op beroepskwalificatie, terwijl er ook andere centrale functies van onderwijs in het geding zijn, waarbij het gaat om doelen als socialisatie en culturele en maatschappelijke vorming. Bij het vervangen van overheidsfinanciering door consumentenbijdragen zouden deze functies in de knel kunnen komen. Ook het ideaal van gelijke kansen in het onderwijs staat op gespannen voet met het Angelsaksische model (vgl. Scheerens, Korevaar & De Rijcke, 1991). Op de verdere afweging van pro's en contra's van bepaalde gradaties van autonomisering zal verderop worden ingegaan (zie ook Dercksen et al. 1993, hfst. 4).

INTERNATIONALE VERGELIJKING

In het kader van het OECD/CERI project, waarin internationale onderwijsindicatoren worden vastgesteld, wordt onder meer een poging gedaan om de zeggenschap van verschillende besluitvormingsniveaus over verschillende beleidsterreinen in kaart te brengen. De uitkomsten hiervan, welke vooralsnog beperkt zijn tot het primair en algemeen voortgezet onderwijs, bieden de mogelijkheid de mate van autonomie van Nederlandse scholen te vergelijken met de situatie in andere OECD-landen.

Het conceptuele kader van de desbetreffende indicatoren bestaat uit 4 besluitvormingsniveaus, 4 categorieën van beslissingen, c.q. beleidsterreinen en 4 zogenoemde "modes of decision making".

De besluitvormingsniveaus die worden onderscheiden zijn de school, een intermediair niveau direct boven de school (in de meeste gevallen is dat de gemeente, maar ook de Amerikaanse school districts en de Engelse Local Education Authorities vallen daaronder), een hoger intermediair niveau (provincies, kantons, staten) en de nationale overheid als het hoogste niveau.

De 4 beleidsterreinen zijn respectievelijk de *inrichting en organisatie van het onderwijs* (beslissingen over methoden, groeperingsvormen, toetsing van leerresultaten en de toelating van leerlingen), *plannings- en structuurbeslissingen* (de stichting of sluiting van een school, het toevoegen of verwijderen van een leerjaar, de keuze van het programma van vakken, de vaststelling van het lesrooster en het vaststellen van exameneisen), *personeelsmanagement* (onder meer aanstelling en ontslag van onderwijzend personeel, arbeidsvoorwaarden, salarisniveaus en de zeggenschap over het carrière-verloop van de staf en schoolleiding) en *de toewijzing en benutting van "resources"* (wie bepaalt het budget van de school).

De modi van besluitvorming hebben betrekking op de vraag of een beslissing volstrekt autonoom door een bepaald niveau wordt genomen, dan wel in samenspraak met een ander niveau, of binnen een formeel kader dat op een ander niveau is vastgesteld. Deze laatste categorie, de modi van besluitvorming, zal hier verder buiten beschouwing worden gelaten, voor nadere informatie zij verwezen naar OECD, 1992, 1993.

De gegevensverzameling vond plaats aan de hand van een gestructureerde vragenlijst (een Engelstalige en Franstalige versie), die werd ingevuld door ter zake deskundige respondenten.

Tabel 2. Besluitvormingsniveaus in 14 OECD-landen (openbaar onderwijs). Bron: OECD, 1994 (in voorbereiding).

	N	School	Lagere intermediair	Hogere intermediair	Land
ISCED 1	14	35	35	10	20
ISCED 2	14	38	31	12	19
ISCED 3	14	41	19	21	19

In de meeste gevallen ging het hierbij om ambtenaren van departementen van onderwijs. De voorgeschreven procedure bestond uit het onafhankelijk van elkaar laten invullen van de vragenlijst door twee teams van 5 respondenten. Op onderdelen waar de scoring tussen de twee teams verschilde vond overleg plaats. Nadat consensus was bereikt werd de uiteindelijke invulling per land geratificeerd door de nationaal coördinator van het OECD indicatorenproject. Opgemerkt moet worden dat deze procedure zich eerder voltrok onder ambtelijke controle dan dat er een onderzoekstechnische monitoring van het scoringsproces plaatsvond.

Er is geen beschrijving gevraagd waaruit afgeleid zou kunnen worden of de procedure naar de letter is toegepast in alle deelnemende landen.

In Tabel 2 wordt een overzicht gegeven van de verdeling van beslissingen over de vier niveaus, gemiddeld over 14 OECD landen, waar Nederland overigens geen deel van uitmaakt.

De tabel geeft het percentage beslissingen aan dat, gemiddeld over 14 landen, op een bepaald niveau genomen wordt. De ISCED (International Standard Classification for Education) -niveaus 1, 2 en 3 staan respectievelijk voor lager onderwijs (dus exclusief kleuteronderwijs), eerste fase voortgezet onderwijs en tweede fase voortgezet onderwijs.

Uit dit overzicht blijkt dat verreweg de meeste beslissingen genomen worden door de scholen en het daar direct bovengelegen bestuurlijke niveau, b.v. de gemeente. Tevens wordt zo'n 20 % van de beslissingen genomen door de nationale overheden. Samengenomen over de veertien landen (te weten: Duitsland, Oostenrijk, België, Denemarken, Spanje, De Verenigde Staten, Finland, Frankrijk, Ierland, Noorwegen, Nieuw Zeeland, Portugal, Zweden en Zwitserland) is het hogere intermediaire niveau, b.v. de provincie van wat minder betekenis. Verder is er een lichte trend zichtbaar van een toename van de schoolautonomie, naarmate het gaat om een hoger schooltype.

In Tabel 3 is de verdeling per land van de beslissingen over de vier niveaus aangegeven, voor de eerste fase van het (openbaar) voortgezet onderwijs. De gegevens voor Nederland zijn ontleend aan een eerdere OECD-publicatie (OECD, 1992) dan de gegevens voor de overige landen, die ontleend zijn aan OECD, 1993). Als peildatum voor beide publicaties wordt overigens 1991 vermeld.

Op basis van de gegevens in de tabel kunnen de landen op verschillende wijze ingedeeld worden. Om te beginnen kan onderscheid gemaakt worden naar het belang van het nationale niveau. Daarbij valt op dat dit alleen in Portugal substantieel is (daar wordt 55% van de beslissingen aangaande de eerste fase van het voortgezet onderwijs door de nationale overheid genomen). Vervolgens is er een groep landen waar de invloed van de nationale overheid juist nihil of beperkt is. In deze groep vallen respectievelijk De Verenigde Staten, Zwitserland en België en Duitsland en Zweden. Een tussenpositie wordt ingenomen door Oostenrijk, Denemarken, Spanje, Finland, Frankrijk, Ierland, Noorwegen, Nieuw Zeeland en Nederland.

Wanneer de invloed van de school op de besluitvorming als uitgangspunt wordt genomen dan kan de volgende volgorde worden opgesteld:

Tabel 3. Verdeling per land van de beslissingen over vier niveaus voor de eerste fase van het (openbaar) voortgezet onderwijs. Bron: OECD, 1992, 1993.

	Besluitvormingsniveau				
%	School-niveau	Intermediair niveau 1	Intermediair niveau 2	Landelijk niveau	Totaal
Oostenrijk	44	8	26	23	100
België	26	50	24	-	100
Denemarken	39	48	-	14	100
Finland	38	50	-	13	100
Frankrijk	35	-	35	30	100
Duitsland	32	44	17	7	100
Ierland	74	8	-	18	100
Nederland	46	28	-	26	100
Nieuw Zeeland	73	-	-	27	100
Noorwegen	31	45	-	24	100
Portugal	42	-	3	55	100
Spanje	28	26	14	32	100
Zweden	47	47	-	6	100
Zwitserland	9	44	46	-	100
Verenigde Staten	26	71	3	-	100

1) Ierland en Nieuw Zeeland (74% en 73%)
2) Zweden en Nederland (47 % en 46%)
3) Oostenrijk, Denemarken, Finland en Portugal (38 – 44%)
4) Duitsland, Frankrijk en Noorwegen 31 – 35%)
5) België, Spanje en de Verenigde Staten (26 – 28%)
6) Zwitserland

Tenslotte is een indeling mogelijk waarbij gekeken wordt naar de mate waarin een bepaald niveau dominant is, dan wel sprake is van meer gedeelde zeggenschap. Dit is weergegeven in het onderstaande overzicht, ontleend aan OECD, 1994.

In Tabel 5 wordt aangegeven hoe de beslissingen die de school neemt verdeeld zijn over de vier onderscheiden beleidsterreinen. Opnieuw zijn de gegevens voor Nederland ontleend aan OECD, 1992, terwijl die voor de overige landen ontleend zijn aan OECD, 1993.

Tabel 4. Onderwijsstelsels naar dominant niveau bij de besluitvorming, eerste fase v.o.

A. Stelsels waar een niveau overheerst:
- school: Ierland, Nieuw Zeeland
- lokaal: België, Verenigde Staten
- regionaal: Zwitserland
- nationaal: Portugal

B. Stelsels met twee-niveau dominante structuren
- school en lokaal: Duitsland, Denemarken, Finland, Noorwegen, Zweden

C. Stelsels met multi-niveau dominante structuren
Oostenrijk, Spanje, Frankrijk en Nederland

Tabel 5. Beslissingen op schoolniveau per beleidsterrein als percentage van alle beslissingen, eerste fase v.o. Bron: OECD, 1992, 1993.

	Beslissingsdomein			
%	Instructie-organisatie	Structuren	Personeels-management	Resources
Oostenrijk	25	8	-	11
België	18	6	-	2
Denemarken	22	9	5	4
Finland	16	13	6	4
Frankrijk	20	3	4	7
Duitsland	22	6	-	4
Ierland	22	18	17	17
Nederland	26	20	-	-
Nieuw Zeeland	25	13	25	10
Noorwegen	18	6	-	7
Portugal	19	6	10	7
Spanje	22	-	2	4
Zweden	22	9	8	7
Zwitserland	9	-	-	-
Verenigde Staten	19	5	2	-

Het algemene patroon dat naar voren komt uit Tabel 5 is het gegeven dat de beleidsruimte van scholen vooral gelegen is binnen het domein van de beslissingen over de inrichting en organisatie van het instructieproces. Slechts in een beperkt aantal landen (Finland, Ierland, Nieuw Zeeland en Nederland) hebben scholen tevens veel zeggenschap over planning en structurering. De zeer hoge score van Nederland op dit domein wordt bepaald door de vrijheid van onderwijs op het terrein van de curriculumplanning. Wat de positie van Nederland op de terreinen van de aanwending van het budget en de arbeidsvoorwaarden betreft doet de score enigszins gedateerd aan. De in eerdere paragrafen geschetste ontwikkeling is duidelijk in de richting van meer zeggenschap van de scholen op deze gebieden.

Over het algemeen geldt verder dat in de landen waar relatief veel beslissingsbevoegdheid op het schoolniveau ligt deze zich uitstrekt over alle vier de beleidsdomeinen. Dit geldt voor Ierland, Nieuw Zeeland, Zweden, Denemarken, Finland en Portugal. Nederland en Oostenrijk vormen hierop een uitzondering.

Aan deze internationaal vergelijkende gegevens kan in de eerste plaats de algemene conclusie worden verbonden dat er in de traditionele patronen van centralistische vs decentralistische onderwijsstelsels veel veranderd is. Zo is er in traditioneel centralistische onderwijsstelsels als Frankrijk, maar vooral ook Zweden behoorlijk gedecentraliseerd. In sommige landen, zoals Spanje en België heeft het regionale niveau sterk aan betekenis gewonnen. Ook is er het patroon van landen waar de traditioneel belangrijke intermediaire niveaus aan importantie inboeten, ten faveure van het school- en het nationale niveau (Het Verenigd Koninkrijk, de Verenigde Staten en Nieuw Zeeland). Zo gaat in Groot Brittannië een streven naar versterking van het schoolmanagement samen met de invoering van nationale curricula en een vermindering van de betekenis van de local educational authorities.

In de tweede plaats blijkt uit de gepresenteerde gegevens dat er in Nederland op het terrein van de autonomie van scholen op het vlak van de curriculumplanning en de inrichting en organisatie van het instructieproces – althans binnen de beperkingen van de in het OECD-project gekozen operationalisaties en uiteraard alleen voor de onderwijssectoren waarop de

gegevens betrekking hebben – weinig meer te winnen is. Wel is duidelijk dat er onderwijsstelsels zijn waarop scholen ook op de gebieden personeelsbeleid en controle over budgetten meer beleidsvrijheid hebben.

Tenslotte moet gewezen worden op enkele methodologische beperkingen van het onderzoek waarop de OECD-indicatoren zijn gebaseerd. In vrijwel ieder land is er sprake van bepaalde specifieke omstandigheden die een eenduidige scoring in de gegeven rubrieken moeilijk maken. In zulke gevallen moesten de respondenten een keuze doen, gebaseerd op de meest voorkomende situatie. Zo bleek het voor Nederland moeilijk om een keuze te maken voor het schoolbestuur of de gemeente, waaronder slechts een school ressorteert: moest dit gerekend worden tot de school of tot het naastgelegen intermediaire niveau? Verder is de afbakening tussen de beslissingsbevoegdheden van wat in Nederland bekend staat als het bevoegd gezag van de school en het schoolmanagement in vele landen niet formeel vastgelegd, zodat de respondenten-teams zich moesten bedienen van een inschatting van de in de praktijk meest voorkomende invulling.

AUTONOMIE EN KWALITEIT

Aan de hoge plaats die deregulering en autonomievergroting momenteel op de onderwijsbeleidsagenda hebben, liggen meerdere motieven ten grondslag. Als belangrijke achtergrond kan de in de loop van de jaren tachtig veranderde "Zeitgeist" genoemd worden, tot uitdrukking komend in een kritischer houding tegenover de staat en het overheidsbeleid als zodanig en een toegenomen preoccupatie met "de markt" als een doelmatiger geacht maatschappelijk sturingsmechanisme. Naast een uitvoeriger beschrijving van deze meer algemene achtergrond, noemt Leune, 1994, een viertal meer directe motieven: het geloof dat autonomere scholen beter onderwijs bieden, de uitbouw van de school als professionele organisatie, waarvoor de nodige speelruimte zou moeten zijn, kostenbeheersing en de behoefte aan "meer en nieuwe machts- en statusposities bij de elite in het schoolwezen".

Met enige toelichting zijn de drie eerste motieven terug te brengen tot één motief, namelijk het motief om te komen tot een grotere doelmatigheid van het onderwijssysteem, door een beter, "bedrijfsmatiger" functioneren van scholen. Doelmatigheidswinst kan dan zowel worden geboekt door kwalitatief beter onderwijs tegen dezelfde of zelfs afgenomen kosten, als door onderwijs van gelijkblijvende kwaliteit bij afgenomen kosten. Organisatieverandering past alleen in dit verhaal, voor zover de aanname gerechtvaardigd is dat "professionelere" organisaties bijdragen tot een verhoogde kwaliteit. Een alternatieve these is dat organisatieverandering een middel is dat andere doelen dient, zoals het tegemoetkomen aan de eisen van flankerend beleid, te weten schaalvergroting, of de legitimering vormt voor het vierde door Leune genoemde motief.

Om nader in te kunnen gaan op de plausibiliteit van de stelling dat schoolautonomie de kwaliteit van het onderwijs bevordert, dient uiteraard eerst duidelijk te moeten worden gemaakt op welke wijze, aan de hand van welke operationele criteria, onderwijskwaliteit gedefinieerd moet worden. Vervolgens kan gekeken worden naar beschikbare theorie die het veronderstelde verband zou kunnen verklaren en tenslotte zou stilgestaan kunnen worden bij de beschikbare empirische evidentie. Deze drie zaken zullen in de volgende subparagrafen kort aan de orde worden gesteld. Daaraan voorafgaand moet echter nog gewezen worden op twee complicerende factoren. De eerste is dat aan het onderwijssysteem als geheel eisen worden gesteld die niet volledig herleidbaar zijn tot het doelmatig functioneren van afzonderlijke scholen. De belangrijkste daarvan is dat coördinatie tussen scholen in de vorm van doorstromingsmogelijkheden van leerlingen gegarandeerd moet zijn. De noodzakelijke waarborging hiervan betekent een belangrijke begrenzing van schoolautonomie, met name in het funderend onderwijs (vgl. Leune, 1994). Een tweede complicerende factor zou kunnen zijn dat de implementatie van de organisatieverandering op schoolniveau zoveel tijd en energie kost dat aan de verbetering van het primaire proces niet wordt toegekomen (Scheerens, 1993).

Onderwijskwaliteitscriteria

Kwaliteit van onderwijs moet blijken uit de onderwijsopbrengsten (De Groot, 1985). In deze typering wordt kwaliteit dus opgevat in termen van output, zodat bijvoorbeeld "de kwaliteit van de organisatie" op zichzelf niet als kwaliteitscriterium wordt gezien. Hoogstens kunnen dergelijke "proces"- of throughput- criteria worden opgevat als secundaire, of ondersteunende criteria (vgl. Scheerens, 1992). Op deze wijze wordt de kwaliteit van de organisatie afgemeten aan de mate waarin de organisatie bijdraagt aan de onderwijsopbrengsten en niet aan een of andere intrinsieke waarde van de organisatiekenmerken zelf. Uiteraard staat of valt deze zienswijze met de aanwezigheid van een causale relatie tussen organisatie en een zodanige inrichting van het primaire proces dat de hoogst mogelijke opbrengsten worden bereikt. Voor zover die relatie niet aantoonbaar aanwezig is zou "spaarzaamheid" het criterium moeten zijn om de schoolorganisatie te beoordelen. (Dit tegen de achtergrond van verschijnselen als goal-displacement en de neiging tot overproduktie van beleid in organisaties binnen de publieke sector).

Voor het hier aan de orde zijnde thema zijn deze nuances op z'n plaats, omdat zowel autonomievergroting, als het flankerende beleid van de schaalvergroting leiden tot een behoefte aan meer management. Gezien een mogelijke overschatting van de betekenis van onderwijsmanagement voor het primaire proces en de incentives die er zijn voor managers van grotere eenheden, waar Leune op wijst, is het niet onwaarschijnlijk dat bij deze ontwikkelingen het doel voorbijgeschoten wordt en er meer management wordt gecreëerd dan functioneel is.

Tot zover het onderscheid tussen primaire en secundaire criteria bij de definiëring van onderwijskwaliteit. Vervolgens is een nadere typering van "onderwijsopbrengsten" gewenst. Hierbij kan een onderscheid gemaakt worden tussen criteria die bepaald worden direct aan het eind van een bepaalde opleiding, zoals een eindtoets of een eindexamen en criteria, die betrekking hebben op een, ook in de tijd, verderweggelegen toepassings-situatie. Eventueel kunnen deze twee typen opbrengstcriteria worden betiteld als leereffecten en transfer-effecten. Zo zou het schoolsucces van leerlingen, gemeten in het voortgezet onderwijs kunnen worden beschouwd als het transfer-effect van het onderwijs op de basisschool. In de situatie van het beroepsonderwijs is het iets meer gebruikelijk om dit onderscheid te maken; transfer-effecten hebben daar het karakter van toepassing in de werksituatie. De twee aspecten die onderscheiden kunnen worden aan de kwaliteitsvraag , namelijk het juiste op de juiste wijze doen ("doing the right things" en "doing things right"), zijn met deze twee typen effecten te verduidelijken. De vraag welke effecten er aan het eind van de opleiding gemeten moeten worden, zou namelijk voor een belangrijk deel beantwoord moeten aan de hand van de wenselijk geachte transfer-effecten.

Verder zijn er nog verschillende classificaties en taxonomieën in omloop om onderwijsopbrengsten nader te typeren; bijvoorbeeld het onderscheid in cognitief, affectief en psychomotorisch en, binnen het cognitieve domein, het onderscheid tussen het stimuleren van cognities resp. meta-cognities (b.v. leren leren), (vgl. Scheerens, 1994, Creemers, 1994). Op de technische problemen bij het meten van de diverse soorten en typen onderwijsopbrengsten, alsmede de moeilijkheid om er zeker van te kunnen zijn dat gemeten effecten inderdaad aan de opleiding kunnen worden toegeschreven (onder meer door middel van *value added* performance measures) wordt hier niet ingegaan.

Met het bovenstaande voor ogen wordt het mogelijk om wat meer duidelijkheid te scheppen omtrent de vraag om welke aspecten van kwaliteit het nu precies gaat bij de meer of minder expliciete verwachting dat autonomie van scholen tot meer kwaliteit leidt.

In een eerdere paragraaf is al gewezen op het feit dat de kwaliteit van het onderwijsstelsel (het systeemniveau dus) niet volledig te herleiden is tot de kwaliteit van het bedrijfsmatig functioneren van scholen. Wat betreft autonomie als middel om het laatste te bewerkstelligen, staat dit streven, in verband met de coördinatie-behoefte binnen het stelsel als geheel, zelfs op gespannen voet met kwaliteitseisen op systeem-niveau.

In sommige publicaties (vgl. Van Wieringen, 1994) wordt kwaliteit in relatie tot meer autonomie van scholen opgevat in de zin van intrinsieke kwaliteitseisen die aan de organisatie gesteld worden en die hierboven zijn opgevat als secundaire kwaliteitscriteria. In zogenoemde

schooleffectiviteitsmodellen (Scheerens, 1994; Creemers, 1994) zijn zowel secundaire als primaire kwaliteitscriteria in het geding. De eerste categorie wordt daarbij dan vooral beschouwd voor zover ze instrumenteel wordt geacht ten opzichte van de tweede categorie. Een voorbeeld hiervan is het construct "onderwijskundig leiderschap", als zijnde een vorm van schoolmanagement die vrij direct is gericht op verbetering van onderwijskundige condities in de school, die op hun beurt weer bijdragen aan de optimalisering van leereffecten. Op de plaats van "autonomie" binnen deze modellen zal verderop nog worden ingegaan.

In het beroepsonderwijs heeft kwaliteit, in de zin van transfer-effecten, een ander karakter dan in het funderend onderwijs. Vergroting van schoolautonomie wordt hier vaak als een middel genoemd om de aansluiting bij behoeften op de arbeidsmarkt te verbeteren. Autonomie wordt daarbij dan gezien als een belangrijke voorwaarde om het aanbod aan opleidingen zo te kiezen dat op het raakvlak van opleidingsbehoeften van studenten en behoeften op de arbeidsmarkt "maatwerk" kan worden geleverd. Daarbij staan dan bijvoorbeeld werkvormen als "co-makership" met het lokale bedrijfsleven voor ogen. Opgemerkt moet worden dat schoolautonomie, in de zin van onderwijskundige produktautonomie, zoals onderscheiden door Leune, in deze alleen noodzakelijk is wanneer dit gebeurt in het perspectief van het eerder aangehaalde Angelsaksische model. Indien, zoals in het continentale model, de bepaling van eindtermen voor het beroepsonderwijs, op branche-niveau plaatsvindt, onder toezicht van de rijksoverheid, is een dergelijke vergaande schoolautonomie niet nodig. In het onderstaande schema worden de meest in aanmerking komende aspecten van autonomie in verband gebracht met verschillende soorten kwaliteit voor de verschillende onderwijssectoren.

Theorie

Wat is er aan theorie beschikbaar ter verklaring van de veronderstelde gunstige invloed van autonomievergroting van scholen? Met andere woorden, welk achterliggend mechanisme zou ten grondslag kunnen liggen aan de veronderstelling dat automomere scholen effectiever en responsiever zouden zijn?

Een eerste aangrijpingspunt zou gezocht kunnen worden in conceptuele schooleffectiviteitsmodellen. Hoewel dergelijke modellen niet de status van theorieën hebben, is in ieder geval wel een poging gedaan om herhaaldelijk in empirisch onderzoek vastgestelde verbanden tussen school-en klaskenmerken enerzijds en outputcriteria anderzijds in onderling verband te plaatsen (Scheerens, 1992; Creemers, 1991; Bosker en Scheerens, 1994). Duidelijk is dat "autonomie" in

	Belangrijkste aspecten van autonomievergroting	Belangrijkste aspecten van autonomieverkleining	Meest beschouwde kwaliteitscriteria
bo/vo	formatie budgetsysteem	eindtermen	leereffecten, secundaire criteria
mbo	- lump sum bekostiging - keuze van opleidingen - mogelijkheid contract-onderwijs	- rendementsbekostiging - eindtermen - toegenomen invloed sociale partners	- leereffecten, secundaire criteria - transfercriteria (als planningsinstrument voor eindtermen)
hbo/wo	- keuze van opleidingen - mogelijkheid contract-onderwijs - eigen CAO		performance indicatoren

Schema 1. Aspecten van autonomievergroting/-verkleining en meest beschouwde kwaliteitscriteria per sector.

deze modellen tot nu toe geen factor is. Als achterliggend mechanisme ter verklaring van wat er in de zin van het schooleffectiviteisonderzoek "werkt" in het onderwijs wordt eerder gedacht aan het cybernetisch principe, dat wil zeggen het functioneren van evaluatie en terugkoppeling, dan aan fundamentele wijzigingen in de beslissingsbevoegdheden op verschillende organisatieniveaus.

Dichter in de buurt komt de zogenoemde "public choice" theorie met het propageren van marktmechanismen in de publieke sector. De institutionele poot van deze theorie (Niskanen, 1971) stelt dat de incentive-structuur van organisaties in de publieke sector zodanig is dat er een neiging ontstaat tot overproduktie van diensten, terwijl er tevens ruimschoots aanleidingen zijn voor managers en medewerkers om af te dwalen van de eigenlijke primaire taakstelling van de organisatie (goal displacement). De remedie die deze theorie levert is het omvormen van openbare instituties in de richting van het model van bedrijven die op de vrije markt opereren. Autonomievergroting en privatisering van scholen passen geheel in deze remedie. Voor de onderwijssector is deze gedachte het verst uitgewerkt door Chubb & Moe (1990). Zij bepleiten een herstructurering van onderwijsstelsels conform wat hier eerder is aangeduid als het "Angelsaksische model". Vrije keuze van scholen door ouders en leerlingen, evenals keuzevrijheid van scholen wat betreft de instroom van leerlingen staan hierin centraal, terwijl de scholen eveneens vrij zijn in het bepalen van het onderwijsaanbod en de kwalificatie-structuur voor het onderwijzend personeel. Opmerkelijk is dat voor Chubb & Moe de keuze tussen overheidsregulering van het onderwijs en "vrije keuze" radicaal is; zij zien niets in mengvormen, waarbij bijvoorbeeld de proceskant geliberaliseerd wordt, maar de produktkant, door accountability-eisen, juist strakker wordt voorgeschreven. Naar hun mening leiden dergelijke mengvormen uiteindelijk toch weer tot verstikkende bureaucratisering. De eerder genoemde bezwaren tegen het Angelsaksische model zijn onverkort van toepassing op de uitwerking die Chubb & Moe geven van de public choice theorie toegepast op het onderwijs. In het bijzonder geldt dit voor het funderend onderwijs. Door critici is verder aangevoerd dat marktmechanismen in het onderwijs imperfect werken, bijvoorbeeld omdat ouders zich door andere dan kwaliteitsoverwegingen laten leiden bij de keuze van scholen voor hun kinderen en omdat het in de praktijk niet gemakkelijk zou zijn om na een gebleken verkeerde schoolkeuze van school te veranderen (Riley, 1990, 558).

Voor zover er een theorie is die kan verklaren waarom vergrote schoolautonomie zou leiden tot meer onderwijskwaliteit is deze, op conceptueel vlak, dus omstreden. In hoeverre de desbetreffende theorie gestaafd wordt door empirische evidentie wordt in de volgende sub-paragraaf bekeken.

Empirische evidentie

Zoals reeds gesteld heeft het schooleffectiviteitsonderzoek tot nu toe geen aanwijzingen opgeleverd dat autonomere scholen effectiever zijn. (Het onderzoek van Chubb & Moe, dat claimt dit uitdrukkelijk wel te doen, zal hieronder nader worden geanalyseerd). Hoogstens is er enige indicatie dat prestatie-eisen die vanuit de omgeving aan scholen gesteld worden doorwerken in een prestatiegerichte oriëntatie binnen de school en langs die weg bijdragen aan relatief betere prestaties. De voorbeelden van onderzoek waarin dit werd aangetoond hebben echter betrekking op prestatie-eisen vanuit een hoger bestuurlijk niveau, te weten het schooldistrictsniveau, zoals dat bestaat in de Verenigde Staten (Kyle, 1985) en zijn dus allesbehalve voorbeelden van autonomievergroting.

Een lijn van onderzoek die verder relevant is voor de aan de orde zijnde vraagstelling betreffende de relatie tussen schoolautonomie en kwaliteit is het onderzoek naar sectoreffecten; onderzoek dus waarin wordt nagegaan in hoeverre de prestaties van openbare scholen verschillen van bijzondere (Engels: "private") scholen. Zeker wanneer getracht wordt zulke verschillen te verklaren met behulp van kenmerken die ook in het schooleffectiviteitsonderzoek naar voren zijn gekomen kan dit type onderzoek als een specifieke soort schooleffectiviteitsonderzoek gezien worden. De bekendste Amerikaanse onderzoeken op dit terrein zijn gebaseerd op het "High School and Beyond"-onderzoek. De verschillende analyses laten een wisselend patroon

zien van duidelijke sectoreffecten tot twijfels, gezien de geringe grootte van de effecten (vgl. Witte, 1990). De richting van de verschillen is echter steeds in het voordeel van de bijzondere scholen. Gezien de grote verschillen tussen de Amerikaanse private en public schools, zowel qua geografische setting als qua leerlingpopulatie blijft er gerede twijfel of uiteindelijk niet alle verschillen in output aan deze verschillen in uitgangssituatie zouden moeten worden toegeschreven. Ook in het Nederlandse onderzoek op dit terrein worden doorgaans sectoreffecten gevonden die in het voordeel zijn van de bijzondere scholen (vgl. Dijkstra, 1992). Gezien het feit dat Nederlandse bijzondere scholen niet in die mate verschillen van openbare scholen als het geval was in het hierboven aangehaalde Amerikaanse onderzoek, zijn de uitkomsten betekenisvoller. Hypothetische verklaringen die zijn gegeven voor de in Nederland aangetroffen sector-effecten liggen in de sfeer van een grotere overeenkomst in onderwijsvisies tussen betrokkenen en een nauwere betrokkenheid van de gemeenschap van ouders en bestuurders bij de school (vgl. Dronkers, 1992). In een recent onderzoek van Hofman, Hoeben en Guldemond, 1994, wordt steun gevonden voor een verklaring die dichter ligt bij het hier aan de orde zijnde thema. Zij vonden namelijk dat denominatie een indirect effect had op leerlingprestaties, waarbij dit effect gemedieerd werd door de mate waarin sprake was van invloed van interne geledingen op bestuursbeslissingen. Dit is dus een voorbeeld waarin het sectoreffect wordt verklaard door de grotere autonomie van bijzondere scholen, meer specifiek door een grotere mate van mandatering van de beslissingsbevoegdheid naar het niveau van de schoolleiding en de leerkrachten.

In het onderzoek van Chubb & Moe (1990) is in twee stappen gezocht naar het effect van grotere schoolautonomie. In de eerste plaats is het effect onderzocht van een index, waarin diverse schoolkenmerken, die uit het schooleffectiviteitsonderzoek naar voren zijn gekomen waren opgenomen, op leerwinst. In een tweede stap is de samenhang onderzocht tussen diverse aspecten van schoolautonomie enerzijds en de index voor schoolorganisatorische effectiviteit anderzijds. In beide stappen vonden zij substantiële effecten. Ten aanzien van de eerste stap betekende dit dat schoolkenmerken als nadruk op schoolse prestaties, onderwijskundig leiderschap, participerende besluitvorming binnen de school en samenwerking tussen leerkrachten een positieve samenhang vertoonden met leerwinst, in de zin van vooruitgang over een periode van twee jaar. Wat de tweede stap betrof werd een positieve samenhang gevonden tussen de afwezigheid van sterke administratieve controle en afwezigheid van sterke regelgeving op het terrein van personeelsbeleid enerzijds en een hoge score op de index van effectief gebleken schoolorganisatiekenmerken anderzijds. Voor zover autonomie in de zojuist aangegeven zin frequenter voorkwam bij "private" schools impliceerde deze bevinding tevens een verklaring van het sector-effect dat binnen het High School and Beyond onderzoek (het gegevensbestand waar het onderzoek van Chubb & Moe op gebaseerd was) werd gevonden. Op basis van deze bevindingen komen deze auteurs tot de radicale aanbeveling van privatisering van het onderwijs die in de vorige paragraaf al is aangehaald.

Bij het onderzoek van Chubb & Moe zijn echter nogal wat methodologische kanttekeningen te plaatsen. In de eerste plaats kan worden geconstateerd dat zij met hun correlationele analyses in twee stappen minder basis hebben voor causale interpretaties dan bijvoorbeeld het geval is bij het onderzoek van Hofman et al. (1994). Zo biedt hun aanpak geen enkele garantie dat de veronderstelde effecten van autonomie niet geheel wegverklaard worden door schoolorganisatiekenmerken als onderwijskundig leiderschap. In de tweede plaats is de interpretatie van de gevonden verbanden bij Chubb & Moe een opeenstapeling van het kunstmatig uitvergroten van effecten. Dat begint al bij de toepassing van een vergelijking van outliers in elk van de hierboven aangeduide stappen, telkens werden scholen in het bovenste kwartiel van de verdeling vergeleken met scholen op het onderste kwartiel van de afhankelijke variabele.

Witte (1990, p. 19 en 20) noemt daarnaast nog drie andere manipulaties die de, in wezen even kleine effecten als gebruikelijk in het schooleffectiviteitsonderzoek, oppeppen. Dit zijn respectievelijk het disproportionele gewicht van elite-scholen bij de positieve outliers, een transformatie procedure toegepast op de index waarop leerprestaties werden gemeten, die ertoe leidde dat de verschillen in leerwinst tussen aanvankelijk laag- en hoogscorende scholen kunstmatig vergroot werden en het uitdrukken van betrekkelijk kleine effecten in leerwinst in termen van

proporties van jaren leerwinst. Zo wordt een gestandaardiseerd effect van .065 uitgedrukt als een verschil in leerwinst tussen hoog- en laagpresterende scholen van .45 jaar. Dit laatste kon onder meer gebeuren, omdat de leerwinst over de desbetreffende periode van twee jaar op zichzelf nogal bescheiden was.

Resumerend kan worden gesteld dat het empirisch bewijs voor een positieve samenhang tussen autonomie en kwaliteit in de zin van toegevoegde waarde in schoolprestaties nog zeer mager is. Ook op basis van de vergelijking van meer en minder geautonomiseerde nationale onderwijsstelsels is geen evidentie voor dit verband te vinden. Ter illustratie is in figuur 1 het verband tussen de invloed van het schoolniveau op de besluitvorming en gemiddelde prestaties in "reading literacy" afgezet, gebruik makend van gegevens uit het OECD/INES project. Hoewel het aantal waarnemingen te gering is om een correlatie te berekenen geeft het scatter-plot weinig aanleiding om een positieve correlatie te verwachten. Ook uit verdere explorations van de relatie tussen schoolautonomie en andere outputindicatoren, uitgevoerd door Meuret, Prod'hom & Stocker (1995), op basis van de gegevens uit het INES-project lijken er geen correlationele samenhangen geïndiceerd te zijn.

Toch wijzen onderzoeksuitkomsten als die van Hofman et al. en Chubb & Moe in de richting van een gunstige invloed van bestuurlijke en organisatorische autonomie op een effectieve inrichting van de schoolorganisatie. Eventueel zijn deze uitkomsten ook als consistent te beschouwen met bedrijfskundige inzichten omtrent "lerende organisaties" (Senge, 1990). Overigens kan hieraan alleen een pleidooi voor partiële autonomievergroting worden ontleend en geenszins een argumentatie voor een volledige herstructurering in de zin van privatisering.

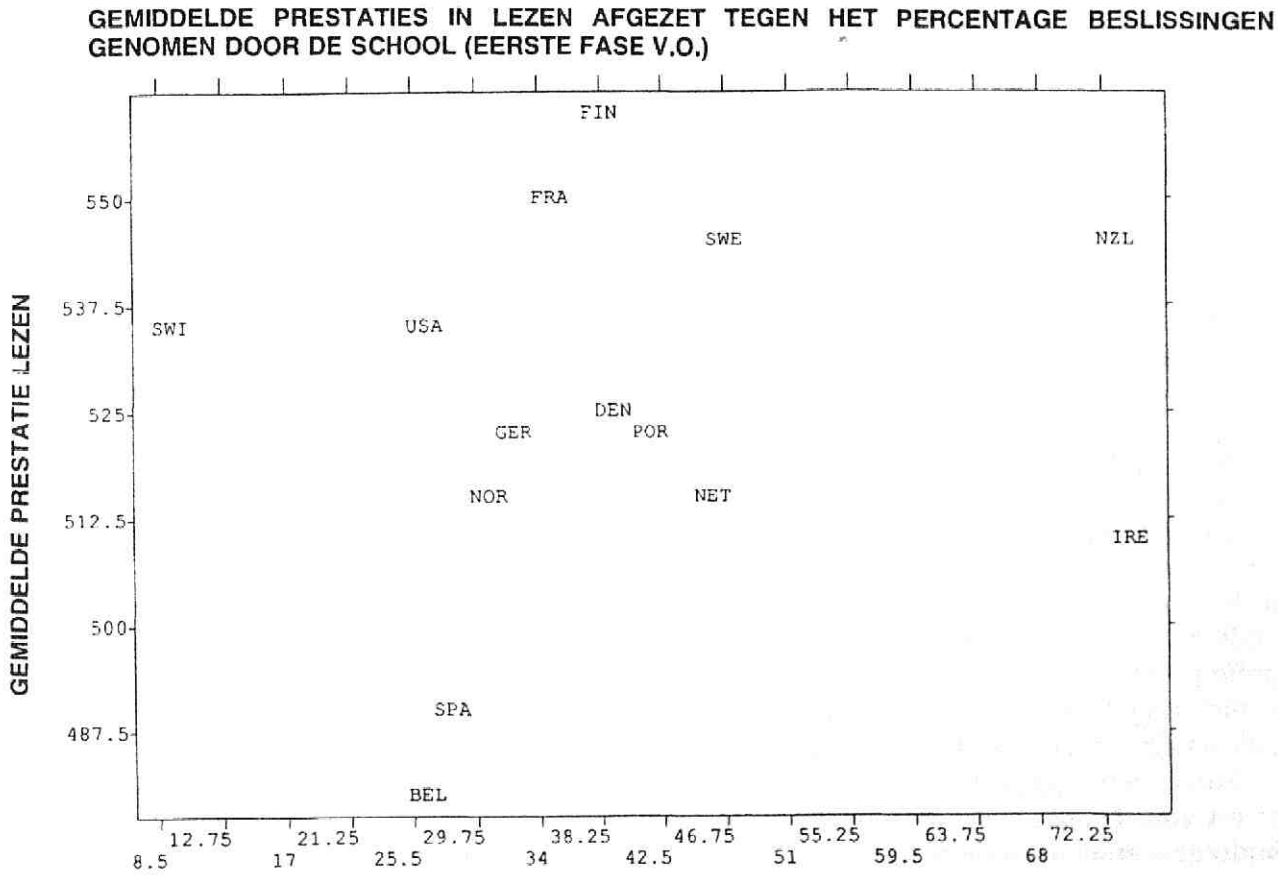


Fig. 1. Percentage beslissingen genomen door de school.

BESLUIT: AUTONOMIE ALS ONDERZOEKSOBJECT

In het voorafgaande zijn nuances geschetst bij de huidige trend om de autonomie van scholen te vergroten. Uit de internationale vergelijking van de OECD blijkt dat, in het primaire en secundaire onderwijs in Nederland, de autonomie van scholen al vrij groot is op pedagogisch-didactisch terrein. Uit de aangehaalde sectorspecifieke invullingen van instellingsautonomie blijkt verder dat voorzichtigheid moet worden betracht met een generaliserende behandeling van het aan de orde zijnde thema over sectoren. Uit de aspectmatige behandeling van het thema autonomie (de verschillende soorten autonomie) blijkt eveneens dat nuancering noodzakelijk is. Vooral op het vlak van de produktmatige onderwijskundige autonomie zijn er belangrijke redenen om deze, in het primaire en secundaire onderwijs, eerder te beperken dan te verruimen (voor een uitvoeriger argumentatie van deze stellingname zij verwezen naar Leune, 1994).

Wat betreft de relatie tussen schoolautonomie en kwaliteit, in de zin van leer- en transfereffecten, kan worden opgemerkt dat de hypothese dat meer autonomie op bestuurlijk en organisatorisch terrein en meer vrijheidsgraden betreffende het personeelsbeleid indirect bijdragen aan de kwaliteit, niet van tafel is. Hoewel de momenteel beschikbare empirische evidentie nog vrij zwak is en de achterliggende theorie nadere toespitsing behoeft, is de gedachte dat meer zelfstandigheid appelleert aan de motivatie en de betrokkenheid van schoolleiders en docenten niet van enige plausibiliteit ontbloot. Zo men wil zou hierbij ook aangevoerd kunnen worden dat bedrijfskundige visies, waarin dit soort principes eveneens benadrukt worden, zoals die van de lerende organisatie, eveneens van toepassing zijn op scholen (al doen deze visies in ongenuanceerde modieusheid niet onder voor de behandelwijze van schoolautonomie welke hier juist bestreden is). De gedachte dat meer autonomie goed is voor de effectiviteit en het innoverend vermogen van scholen is ook een van de peilers van sommige aspecten van wat momenteel in de Verenigde Staten bekend staat als "restructuring". De zo vrijkomende speelruimte voor schoolleiders ("school-based management") en deelname van de leerkrachten aan de besluitvorming binnen de school ("teacher empowerment") zou ten goede komen aan de kwaliteit van het onderwijs. Opmerkelijk is overigens dat daarbij in meerdere publicaties op het belang van een meer centrale planning van curriculum en instructie wordt gewezen (zie bv. Wohlstetter et al., 1994).

Het onderwijsresearchprogramma dat eventueel te verbinden is aan het beleidsprogramma van de vergroting van de autonomie van scholen zou zich met name op de indirecte effectiviteitsverhogende werking van deze bestuurlijk-organisatorische maatregelen moeten richten. Een tweede hoofdthema van dit researchprogramma zou zich moeten richten op de kwaliteitsbewaking en de zelfevaluatie van scholen, in relatie tot meer centralistische eisen aan scholen om verantwoording af te leggen over de produkten van hun dienstverlening. Met name is het hierbij interessant of combinatievormen van evaluatieprocedures kunnen worden ontworpen, die zowel recht doen aan principes van "selfimprovement" als "accountability".

In de respectievelijke bijdragen in dit themanummer wordt een tipje opgelicht van de sluier die ligt over de vraag naar de implicaties van het op autonomievergroting gerichte onderwijsbeleid voor de onderwijspraktijk. De voorbeelden hebben betrekking op opvattingen van scholen in het primair en secundair onderwijs over concurrentie en rendementsbekostiging, schoolleiderschap in het basisonderwijs, en mogelijkheden tot schoolzelfevaluatie in het basisonderwijs. In de bijdrage van Van Amelsvoort en Witziers wordt nagegaan hoe betrokkenen staan tegenover principes als vergroting van de responsiviteit, deregulering en de aanname dat dit tot kwaliteitsverbetering leidt. Verder is een apart deelonderzoek gewijd aan rendementsbekostiging. De uitkomsten bevestigen de sectorspecifieke differentiatie die in het voorafgaande is benadrukt.

Stoel heeft nagegaan hoe de feitelijke tijdsbesteding van schoolleiders in het basisonderwijs er uit ziet. In een meer geautonomiseerd scholenstelsel bestaan hoge verwachtingen van de bijdrage van schoolleiders (vgl. de term school-based management die in dit kader bij de restructuring-beweging in de USA en in het Verenigd Koninkrijk wordt gebezigd). Het onderzoek geeft inzicht in de ruimte die er op dit aspect in het basisonderwijs lijkt te zijn. In de bijdrage

van Bosker en Scheerens wordt ingegaan op enkele meer technische aspecten bij de aspiratie om het evaluerend vermogen van scholen te vergroten. Meer aandacht voor zelfevaluatie van scholen en meer centrale outputcontrole zijn aspecten die, zoals in het voorafgaande gesteld, vaak in één adem met autonomievergroting worden genoemd. Mogelijk ligt hier ook de sleutel voor de verbinding van autonomievergroting en kwaliteit.

Het enigszins fragmentarische karakter van deze thema's is aan de ene kant illustratief voor de veelzijdigheid van het thema organisatie en management van de meer autonome school. Anderzijds weerspiegelt de verbrokkeldheid een gebrek aan programmering van fundamenteelstrategisch onderzoek op het gebied van de derde geldstroom in Nederland. Hopelijk zal de nieuwe organisatievorm voor deze onderzoeksprogrammering, die zijn intrede zal doen na de definitieve expiratie van de Wet op de Verzorgingsstructuur, in dit opzicht beter werken.

LITERATUUR

Bosker, R.J. & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. In R.J. Bosker, B.P.M. Creemers & J. Scheerens (eds.). *Conceptual and methodological advances in educational effectiveness research.* Special issue of the *International Journal of Educational Research.* Oxford: Elsevier Science.

Chubb, J.E. & Moe, T.M. (1990). *Politics, markets and American schools.* Washington, D.C.: Brookings Institute.

Creemers, B.P.M. (1994). De realisatie van de kwaliteit van het onderwijs. In B.P.M. Creemers (red.). *Deregulering en de kwaliteit van het onderwijs.* Groningen: RION.

Dercksen, W.J. & Lieshout, H. van (1993). Beroepswijs Onderwijs. *Ontwikkelingen en dilemma's in de aansluiting van onderwijs en arbeid.* Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.

Dijkstra, A.B. (1992). *De religieuze factor. Onderwijskansen en godsdienst: een vergelijkend onderzoek naar gereformeerd-vrijgemaakte scholen.* Nijmegen: ITS.

Dronkers, J. (1992). Blijvende organisatorische onderwijsverzuiling ondanks secularisering: een onbedoeld effect van overheidsbeleid? *Beleid & Maatschappij, 5,* 227-237.

Groot, A.D. de (1986). Is de kwaliteit van het onderwijs te beoordelen? In A.D. de Groot (red.). *Begrip van evalueren.* Den Haag: VUGA.

Hofman, R., Hoeben, W. & Guldemond, H. (1994). Denominatie en effectiviteit van schoolbesturen. *Tijdschrift voor Onderwijsresearch, 20,* 63-78.

Kyle, M.J. (ed.) (1985). *Reaching for excellence: an effective schools sourcebook.* Washington, D.C.: US Government Printing Office.

Leeuw, A.C.J. de (1982). *Organisaties, management, analyse, ontwerp en verandering.* Assen: Van Gorcum.

Leune, J.M.G. (1994). Onderwijskwaliteit en de autonomie van scholen. In B.P.M. Creemers (red.). *Deregulering en de kwaliteit van het onderwijs.* Groningen: RION.

Meuret, D., Prod'hom, J. & Stocker, E. (1995). L'Approche quantitative dans l'étude des processus de décision au sein des systèmes éducatifs. In J. Scheerens (ed.). *Issues in the development of international process indicators on the functioning of schools and school systems* (provisional title). Paris: OECD.

Ministerie van Onderwijs en Wetenschappen / Schevenings Beraad Bestuurlijke Vernieuwingen (1993). Gezamenlijke richtinggevende uitspraken. Zoetermeer.

Ministerie van Onderwijs en Wetenschappen / Schevenings Beraad Bestuurlijke Vernieuwingen (1994). Gezamenlijke uitspraken van de minister van onderwijs en wetenschappen en de vertegenwoordigers van de koepelorganisaties op grond van de rapportages van de uitwerkingsgroepen. Zoetermeer.

Niskanen, W.A. (1971). *Bureaucracy and representative government.* Chicago: Aldine-Atherton.

OECD/CERI (1992). *Education at a Glance 1.* OECD Indicators. Paris: OECD.

OECD/CERI (1993). *Education at a Glance 2.* OECD Indicators. Paris: OECD.

OECD/CERI (1994). *Decision-making in fourteen educational systems in OECD countries* (in press).

Riley, D.D. (1990). Should market forces control educational decision making? *American Political Science Review, 84,* 554-558.

Scheerens, J. (1992). *Effective schooling, theory, research and practice.* London: Cassell.

Scheerens, J. (1993). *Het onderwijsbeleid op z'n kop?* Voordracht ten behoeve van de studiedag voor medewerkers van het Ministerie van O&W en hoogleraren Onderwijskunde op 1 april 1993. Enschede: Universiteit Twente, TO/O&M.

Scheerens, J. (1994). The school-level context of instructional effectiveness: a comparison between school effectiveness and restructuring models. *Tijdschrift voor Onderwijsresearch, 19*(1), 26-39.

Scheerens, J., Korevaar, G.J. & Rijcke, F.J.M. de (19991). Onderwijsindicatoren in Engeland, oftewel de vele valkuilen van een evaluatie-georiënteerde beleidsvoering opnieuw gellustreerd. *Tijdschrift voor Onderwijswetenschappen, 21*(3/4), 221-227.

Senge, P.M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization.* New York: Doubleday Currency.

Wieringen, A.M.L. van (1994). Deregulering per onderwijssector. In B.P.M. Creemers (red.). *Deregulering en de kwaliteit van het onderwijs.* Groningen: RION.

Witte, J.F. (1990). *Understanding High School Achievement: After a Decade of Research, Do We Have Any Confident Policy Recommendation?* Paper delivered at the 1990 Annual Meeting of the American Political Science Association. San Francisco.

Wohlstetter, P., Smyer, R. & Mohrman, S.A. (1994). *New Boundaries for School-Based Management: The High Involvement Model.* Paper presented at the AERA-convention, New Orleans.

Manuscript ontvangen 28-4-1994
Definitieve versie ontvangen 2-9-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 2, pp. 115-132

Deregulering en de marktpositie van scholen

H.W.C.H. van Amelsvoort en B. Witziers
Universiteit Twente, Enschede

ABSTRACT

This article presents the results of two research projects that were conducted within the framework of the present govermental policy with regard to deregulation and decentralization of decision making power. The first research project dealt with the changes schools go through as a result of this policy and the opinions held by school members about aspects like the legitimacy of the governmental policy and the perceived consequences. The results show that most schools react favorably towards this policy in as much that they increasingly fulfill the conditions necessarily to adjust to the new situation. Considering the opinions held by school members the conclusion is that the govermental policy is favored mostly by members of junior vocational schools and members of school boards and the least by members of primary schools and parents. The second research project investigated the opinions held by principals of secondary schools with regard to the planned introduction of a system of output financing. These results show that most principals oppose the introduction of such a system. However, there are no indications that schools will severely obstruct the implementation of output financing.

INLEIDING

Het onderwijs in Nederland is altijd in sterke mate gebonden geweest aan regelgeving van de kant van de overheid, die dan ook altijd een belangrijke invloed heeft gehad op de vormgeving van het Nederlandse onderwijs. Van Wieringen (1987) en Smêts (1985) spreken in dit verband van een relatie tussen overheid en lokale schoolorganisatie die gekenmerkt wordt door een vanaf de onderwijspacificatie toenemende regulering en centralisering van het onderwijsbeleid. Als oorzaken voor deze sterke wet- en regelgeving worden vaak redenen van kwaliteitsbewaking en financieel beheer genoemd. Daarnaast speelt ook het gelijkheidsprincipe en sterke rol; een sterke wet- en regelgeving betekent dat scholen in hoge mate onder dezelfde condities opereren, waardoor de verschillen tussen scholen gering zullen zijn.

Zoals bekend treedt er pas de laatste jaren in deze situatie een verandering op. Kernwoord hier is deregulering die in beheersmatig en inhoudelijk opzicht tot een van de belangrijkste doelstellingen van het Ministerie van Onderwijs en Wetenschappen in de negentiger jaren behoort (Giesbers, 1990). Onder deregulering wordt in het algemeen het streven verstaan naar een beperkter, globaler en helderder regelgeving (Giesbers, 1987).

Voor het primair en voortgezet onderwijs komt het streven naar deregulering tot uiting in de ministeriële nota "De school op weg naar 2000" . Uitgangspunt van deze nota is het toekomstig beeld van de autonome school. Naast het beperken van de overheidsuitgaven heeft deregulering namelijk tot doel de autonomie van scholen te vergroten. Daarmee zou met name de doelmatigheid van scholen vergroot kunnen worden, evenals de flexibiliteit waarmee ze op veranderingen in onderwijsbehoeften kunnen inspelen. Met andere woorden, door meer zelfstandigheid zullen scholen beter toegerust zijn om aansluiting te vinden bij een maatschappelijke omgeving die gekenmerkt wordt door snelle veranderingen en een toenemende diversificatie op lokaal en regionaal niveau (Toenders, 1990). Dat juist op het punt van aansluiting tussen school en omgeving het onderwijs tekort schiet, blijkt uit de kritiek hierover. Zo wordt onder andere gewezen op de gebrekkige aansluiting tussen onderwijs en arbeidsmarkt (Van Hoof, 1986; Geurts, 1989) en het verwaarlozen van de belangen van onderwijsconsumenten (Versloot, 1990;

* Adres: Faculteit Toegepaste Onderwijskunde, Universiteit Twente, Postbus 217, 7500 AE Enschede.

Wansink, 1992). Deze kritiek is overigens een internationaal verschijnsel. Zo heeft in de Verenigde Staten deze kritiek geleid tot onderwijshervormingen die voor een belangrijk gedeelte zijn ontleend aan de "public-choice"-theorie. Deze theorie stelt deregulering en het vergroten van de keuzevrijheid van onderwijsconsumenten voorop met als doel de concurrentie tussen onderwijsinstellingen te stimuleren. De verwachting daarbij is een grotere kwaliteit in de dienstverlening en een meer op de wensen van de onderwijsconsumenten toegesneden aanbod (cf. Boyd & Crowson, 1983; Chubb & Moe, 1990; Lieberman, 1989).

Globaal zijn er binnen een onderwijsinstelling vier gebieden te onderscheiden waar regelgeving vanuit de overheid van toepassing is (Giesbers, 1987; Jacobs, 1987; Marx, 1987):
- regels inzake inhoud en kwaliteit van het onderwijs
- regels inzake personeelsformatie en rechtspositie
- regels inzake huisvesting, materiële voorzieningen en bekostiging
- regels inzake de schoolorganisatie en medezeggenschap

Door deze regelingen van de overheid wordt de beleidsruimte van scholen begrensd. De ruimte die overblijft voor scholen om eigen beleid te voeren wordt vaak uitgedrukt als de relatieve autonomie van een school (Giesbers, 1987). Door deregulering zal deze relatieve autonomie toenemen. De belangrijkste consequentie daarvan is dat scholen gestimuleerd worden eigen (bestuurlijk) beleid te voeren.

Uit de genoemde nota "De school op weg naar 2000" blijkt dat de beslissingsvrijheid van scholen vooral zal groeien ten aanzien van inzet van personele en financiële middelen en de realisering van een maximaal onderwijsresultaat, gericht op het behalen van de door de overheid vastgestelde kerndoelen (Giesbers, 1990; Toenders, 1989). De school zal daartoe door middel van invoering van het formatie-budgetsysteem een naar eigen inzicht te besteden budget krijgen. Daarmee kan ze zelf de personeelsformatie en de daarin gewenste functies vastleggen, waardoor functie- en taakdifferentiatie tot de mogelijkheden gaan behoren evenals aandacht voor loopbaanontwikkeling. Verder krijgen scholen de mogelijkheid extra financiën te verwerven door onderwijs aan derden te leveren tegen betaling. Tenslotte krijgen scholen in het voortgezet onderwijs door de introductie van de basisvorming meer keuzevrijheid wat betreft hun onderwijsaanbod. De overheid zal bij dit alles de kwaliteit bewaken door middel van toetsing achteraf, waarbij geformuleerde eindtermen een belangrijke rol zullen spelen als maatstaf (Giesbers, 1990). Overige instrumenten van kwaliteitszorg in dit kader zijn de eisen voor aanwezigheid van een systeem voor kwaliteitszorg en de verplichting van een jaarverslag.

In algemene zin kan gesteld worden dat met deze maatregelen een voorzichtig begin gemaakt wordt met de introductie van (quasi-) marktmechanismen in het onderwijs. Een gevolg van deregulering is immers dat concurrentie tussen scholen zal toenemen, vooral vanwege het feit dat de verschillen tussen scholen zullen toenemen. Dat de overheid nog verder wil gaan met de introductie van marktmechanismen moge blijken uit de discussies omtrent de wenselijkheid van bekostiging op basis van rendement en de openbaarmaking van schoolresultaten (Amelsvoort, Bosker & Brandsma, 1992). Als assumpties gelden bij deze vormen van outputcontrole dat scholen de kwaliteit van het onderwijs daadwerkelijk kunnen beïnvloeden en dat deze maatregelen scholen zullen stimuleren kwaliteitgericht te werken. Introductie van deze maatregelen zou het begin zijn van werkelijke marktmechanismen, met name omdat scholen voor een belangrijk gedeelte op basis van het gerealiseerde rendement beloond zullen worden. In beperkte mate is overigens een vorm van rendementsbekostiging al ingevoerd in het MBO.

Het proces van deregulering en decentralisatie verkeert nog in een overgangsfase. Vooral in het MBO stuiten de instellingen bij de uitvoering van het beleid op problemen, die dreigen te leiden tot kritiek en weigering tot medewerking (Van Leijenhorst, 1993). De problemen hebben te maken met de beperkte financiële middelen, de tijdelijke regelgeving voor het verplicht in dienst nemen van (dure) wachtgelders en het ervaren contrast tussen de beperkte handelingsvrijheid en de grote verantwoordingsplicht enerzijds en de wens tot autonomievergoting anderzijds (Visscher & Van Amelsvoort, 1993). Hoewel de overheid ernaar streeft om de beleidsruimte van instellingen te vergroten geeft zij zelf ook aan dat deze autonomie nog beperkt is en dat ze

de komende jaren onder meer via nieuwe wetgeving zal werken aan verdere autonomievergroting (Ritzen, 1993).

VRAAGSTELLINGEN

Duidelijk is dat de nieuwe besturingsfilosofie van de overheid scholen voor een aanzienlijke overgang plaatst en een beroep doet op de veranderingscapaciteit van scholen, terwijl deze over het algemeen niet overschat mag worden (Van Gennip, 1991; Commissie Evaluatie Basisonderwijs, 1994). Van uitvoerders van onderwijsbeleid worden scholen in toenemende mate beleidsbepalers. Het is onoverkomelijk dat de toename in bestuurlijke bewegingsvrijheid om een eigen invulling vraagt. De school zal een eigen (bestuurlijk) beleid moeten gaan voeren gericht op haar eigen specifieke situatie.

Het omgaan met autonomie en het creëren van eigen beleid stelt de nodige eisen aan een organisatie. Ten eerste zal de school zich een toekomstgerichte werkwijze moeten aanmeten, tot uiting komend in strategisch management (Van de Berg, 1992; Toenders, 1989). Ze zal zelf in wisselwerking met de omgeving haar onderwijsaanbod moeten bepalen en tevens tijdig kunnen bijstellen. Ook zal ze een meer zelfstandige invulling moeten geven aan haar personeelsbeleid, waarbij afwegingen tussen belangen van personeel enerzijds en belangen van de school en leerlingen anderzijds, centraal zullen staan. Dit alles vereist een goede informatievoorziening over de omgeving, de markt, de kwaliteit van de docenten en het onderwijs.

Ten tweede zal de toename in beleidsvrijheid scholen stimuleren tot een marktgerichte houding (Gielen, 1987; Marx, 1987; Toenders, 1989). Dat wil zeggen dat de school een beleid voert waarin serieus rekening gehouden wordt met de wensen van doelgroepen. De schoolleiding dient daarbij op zijn minst veel aandacht te geven aan openheid van de school ten opzichte van de doelgroepen: de (potentiële) leerlingen en hun ouders, de instellingen die vervolgonderwijs geven, toeleverende scholen en de mogelijke werkgevers van afgestudeerde leerlingen. In concreto betekent dit dat de school een op haar eigen situatie afgestemd beleid dient te voeren, waarmee ze invulling kan geven aan het bieden van een alternatief voor onderwijs aan concurrerende instellingen (Giesbers, 1990). Om te kunnen overleven zal daarbij een marktgerichte houding noodzakelijk zijn, zeker als andere factoren als het overheidsbeleid inzake schaalvergroting, de daling van het aantal leerlingen (S.C.P., 1990) en de veranderende schoolkeuzemotieven van ouders en leerlingen (Boef-van der Meulen, Bronnenman-Helmers, Konings-van der Snoek, 1983; Pelkmans, 1983; Pieters, 1992; Versloot, 1990; De Vries, 1987) in ogenschouw genomen worden.

Uit het bovenstaande kan geconcludeerd worden dat schoolorganisaties in toenemende mate gedwongen worden om in wisselwerking met de omgeving een eigen beleid te voeren. Een belangrijke vraag in dit kader is of en in hoeverre scholen daartoe in staat zijn en welke problemen zij daarbij ondervinden. Wanneer scholen op dergelijke punten onderzocht worden kan enig licht geworpen worden op de vragen of scholen (kunnen) voldoen aan eisen die vanuit het dereguleringsbeleid aan scholen worden gesteld.

Een andere relevante vraag vloeit voort uit het feit dat het beleid inzake deregulering een marktgerichte houding bij scholen wil bevorderen. Als zodanig heeft dit aspect geleid tot de vraag hoe het met deze marktgerichte houding binnen scholen is gesteld. In enge zin kan hier bijvoorbeeld gedacht worden aan de wenselijkheid om het onderwijs af te stemmen op relevante doelgroepen. In bredere zin aan de houding van scholen ten opzichte van (voorgenomen) maatregelen in het kader van deregulering.

Van belang zijn deze houdingen met name vanwege het feit dat in de literatuur veelal naar voren komt dat de implementatie van beleid sterk afhankelijk is van de mate waarin beleidssubjecten zich kunnen vinden in het (voorgestane) beleid (De Graaf & Hoppe, 1991). Dat hier specifiek ten aanzien van het dereguleringsbeleid mogelijk een probleem ligt, kan ontleend worden aan theorieën die het non-responsieve karakter van schoolorganisaties benadrukken (Boyd & Crowson, 1983; Mintzberg, 1979). Vanuit deze optiek is het succes van de implemen-

tatie van een beleid juist gericht op het vergroten van een marktgerichte houding zeker niet vanzelfsprekend.
Samenvattend gaat het om een tweetal vraagstellingen:
a) komen scholen tot een marktgericht beleid zoals bedoeld door het dereguleringsbeleid en welke knelpunten ondervinden zij hierbij ?
b) welke houdingen bestaan er in het onderwijs aangaande de wenselijkheid van het marktgericht functioneren van scholen en een overheidsbeleid gericht op het stimuleren hiervan ?

Deze twee vraagstellingen zullen beantwoord worden aan de hand van een tweetal inhoudelijk aan elkaar gerelateerde onderzoeken. In het eerste onderzoek (SVO-1124) stond de vraag naar het beleid dat scholen voeren om hun marktpositie te verstevigen centraal. Daarnaast werd de houding van betrokkenen bij de onderzochte scholen ten opzichte van marktgerichtheid gemeten, evenals hun houding ten opzichte van een overheidsbeleid gericht op het bevorderen hiervan. Het tweede onderzoek (SVO-0524) betrof een onderzoek naar de houdingen van scholen ten opzichte van rendementsbekostiging in het algemeen voortgezet onderwijs.

HET EERSTE ONDERZOEK: SCHOLEN EN HUN MARKTPOSITIE

Inleiding

Zoals bij de vraagstelling naar voren is gekomen heeft het eerste onderzoek betrekking op de vraag of scholen in staat zijn een beleid te voeren zoals bedoeld door het dereguleringsbeleid van de overheid. Tevens werd de houding van betrokken ten opzichte van de wenselijkheid van markgericht functioneren van scholen gepeild, evenals de houding ten opzichte van het overheidsbeleid gericht op het bevorderen hiervan.

Onderzoeksopzet en steekproef

De onderzoeksopzet van dit onderzoek kan als een vergelijkende case-study getypeerd worden. Bij het onderzoek zijn 17 scholen verdeeld over drie schooltypen (zes scholen in het basisonderwijs, zes scholen in het algemeen voortgezet onderwijs en vijf scholen in het middelbaar beroepsonderwijs). Dit aantal is vooral bepaald door de financiële randvoorwaarden en specifieke eisen die de opdrachtgever aan het onderzoek stelde. In dit opzicht vormde dit aantal voor de onderzoekers een gegeven. Wel binnen de vrijheidsmarges van de onderzoekers lagen de vragen welke sector voor middelbaar onderwijs voor nader onderzoek aan de orde zou komen en uit welke gebieden de scholen geselecteerd zouden worden. Van belang is daarbij de notie geweest dat voor het onderzoek vooral die scholen interessant zouden zijn die zich aantoonbaar in een situatie bevinden waarin marktgerichtheid van belang is. Voor het middelbaar beroepsonderwijs heeft deze overweging er toe geleid dat het onderzoek zich alleen op de sector economie heeft gericht. De prognoses wijzen uit dat juist deze sector sterk te maken heeft met een daling in leerlingaantallen (S.C.P., 1990). Voor de selectie van de te onderzoeken gebieden heeft bovengenoemde notie geleid tot een keuze van een vijftal nodale gebieden[1] waar als gevolg van demografische ontwikkelingen sprake is van een daling in het aanbod van leerlingen. Daartoe is gebruik gemaakt van gegevens van het C.B.S. (1990). Bij de keuze voor deze nodale gebieden speelden naast de demografische ontwikkelingen factoren een rol als een zekere spreiding over Nederland, de urbanisatiegraad, de afstand in kilometers tussen gemeenten binnen deze nodale gebieden en de hoeveelheid scholen die in potentie in aanmerking komen om deel te nemen aan het onderzoek.

Om scholen van hetzelfde schooltype (basisscholen, AVO/VWO, MBO) zo zuiver mogelijk met elkaar te kunnen vergelijken, lag het voor de hand er zoveel mogelijk voor te zorgen dat ze qua externe factoren in vergelijkbare omstandigheden verkeren. Daarom zijn binnen elk gebied een paar grotere gemeenten uitgezocht waarvan meerdere scholen van hetzelfde typen aangeschreven werden om mee te werken aan het onderzoek zodat -demografisch gezien- de context van deze scholen hetzelfde zou zijn. De resultaten van de gevolgde steekproefprocedure staan vermeld in Tabel 1.

Tabel 1. Resultaten van de steekproef.

nodaal a: drie MBO – opleidingen
nodaal b: drie basisscholen en twee scholengemeenschappen
nodaal c: twee basisscholen en een MBO-opleiding
nodaal d: twee MAVO's, een MBO-opleiding een basisschool
nodaal e: een scholengemeenschap

Over ongeveer de helft van deze deelnemende scholen moet tenslotte opgemerkt worden dat het in die zin uitzonderlijke scholen zijn, dat zij ondanks negatieve demografische ontwikkelingen in hun omgeving er in toch in slagen hun leerlingaantal te doen stijgen. Dit is in zoverre een gevolg van de steekproefprocedure dat vooraf op basis van informatie van gemeenten en het C.B.S. (1990) gepoogd is per nodaal die scholen te selecteren waar over de jaren heen sprake is van een stijgende dan wel dalende tendens wat betreft het leerlingaantal. Het doel hiervan was verschillen op het spoor te komen tussen scholen die succesvol zijn in het op peil houden van het leerlingbestand en scholen die in dit opzicht minder succesvol zijn.

Samenvattend kan gezegd worden dat wat betreft regio, schooltype en kenmerken van scholen er steeds gezocht is naar de 'extreme' gevallen. Dat dit (mogelijkerwijs) tot een vertekening van de resultaten heeft geleid ligt voor de hand. Zo zal in een situatie waar scholen veel moeite moeten doen om hun leerlingbestand op peil te houden, men anders handelen en denken.

Instrumentontwikkeling

Zoals hierboven gesteld kan de onderzoeksopzet als een vergelijkende case study worden beschouwd.

Bij de vormgeving van de instrumenten is gebruik gemaakt van de (school)marketing-theorie (Dulmers, 1985; 1987; Faes & Van Tilborgh, 1983; Kotler & Fox, 1985; Mahieu, 1985). In algemene zin houdt deze theorie zich bezig met de relatie schoolorganisatie – omgeving. De veronderstelling is dat naarmate de school meer kennis heeft over de eisen en wensen die leven in de omgeving en deze beter weet te vertalen in het uit te voeren beleid, de marktpositie van de school steviger zal zijn. Meer specifiek geeft deze theorie aan welke middelen een school kan gebruiken om zijn marktpositie te verbeteren. Deze middelen worden veelal onder een aantal categorieën samengevat.

In dit onderzoek is gebruik gemaakt van de indeling van Mahieu (1985). Hij onderscheidt de categorieën produkt (bijvoorbeeld onderwijsaanbod en -aanpak), promotie (bijvoorbeeld reclame en voorlichting), plaats (bijvoorbeeld bereikbaarheid van de school), prijs (bijvoorbeeld ouderbijdrage), profijt (bijvoorbeeld de waarde van het diploma en/of beroepsperspectieven) en personeel (bijvoorbeeld kwaliteit van het personeel).

Deze categorieën hebben in de eerste plaats als leidraad gediend voor de vormgeving voor het interview. Hiervoor is met betrekking tot elke categorie een serie vragen ontwikkeld. Deze vragen hadden als doel te onderzoeken welk beleid scholen op dit punt voerden en welke wijzigingen zich hierbij de afgelopen jaren ten gevolge van de in de inleiding genoemde ontwikkelingen hebben voorgedaan. Daarnaast zijn vragen geformuleerd over de context van de school.

Verder vormden deze categorieën de basis voor een output-lijst. Deze lijst bestond uit een reeks van (mogelijke) activiteiten die een school onderneemt om haar marktpositie te verbeteren. Op deze lijst kon aangegeven worden of men deze activiteiten wel of niet binnen de school ten uitvoer brengt en of hier in vergelijking met een aantal jaren geleden veranderingen waren opgetreden. Daartoe werd gevraagd of men deze activiteiten ook in 1987 ten uitvoer bracht. In dit opzicht is er dus sprake van een retrospectieve meting met mogelijk kwalijke gevolgen voor de betrouwbaarheid vandien.

Om de houdingen van betrokkenen bij de onderzochte scholen te meten is een vragenlijst ontwikkeld bestaande uit oorspronkelijk 43 items. Op basis van een factor-analyse, die als

Tabel 2. Schaalgegevens.

	aantal items	Cronbach's α
responsiviteit	12	0.83
deregulering	16	0.79
kwaliteitsoordeel	4	0.79

heuristische methode gebruikt is om te kijken of er een bepaalde structuur aan de items ten grondslag ligt, een betrouwbaarheidsanalyse en op basis van inhoudelijke gronden zijn uiteindelijk een drietal schalen geconstrueerd. De eerste schaal (responsiviteit) heeft betrekking op de bereidheid om het onderwijs af te stemmen op relevante groepen in de omgeving van de school zoals het bedrijfsleven, ouders en leerlingen. Naarmate de score op deze schaal hoger is, is de bereidheid om rekening te houden met relevante groepen in de omgeving groter. De tweede schaal (deregulering) meet de houding ten opzichte van deregulering, of, met andere woorden, de mate waarin de betrokkenen al dan niet voorstander zijn van een strakke regelgeving aangaande het functioneren van scholen. Een hoge score houdt in dat men het functioneren van scholen zo min mogelijk aan banden gelegd wil zien. De derde schaal (kwaliteitsoordeel) tenslotte geeft inzicht in de verwachte gevolgen van een beleid gericht op het verbeteren van de marktpositie van scholen. Een hoge score op deze schaal betekent dat men de gevolgen voor het functioneren van scholen positief inschat. De relevante gegevens over de ontwikkelde schalen staan vermeld in Tabel 2.

Onderzoeksprocedure

Op elk van de deelnemende scholen werd een diepte-interview gehouden met een leidinggevende. Bij de basisscholen betrof dit altijd de (adjunct-) directeur, op de scholen voor AVO/VWO de rector en op de MBO-opleidingen economie de directeur van de betreffende opleiding of een lid van de centrale directie. Indien tijdens het interview bleek dat er op de school een speciale groep of persoon verantwoordelijk was voor een bepaald aspect van de marktpositie van de school, werd een nader gesprek geregeld. Daarbij werden in ieder geval de vragen voorgelegd die ook in het oorspronkelijke interview met de leidinggevende aan de orde waren geweest. In totaal zijn 25 personen geïnterviewd. Na het gesprek is de outputlijst aan de betrokkenen overhandigd en is gevraagd om relevante documenten.

De vragenlijst betreffende de houdingen is met behulp van de leidinggevende verspreid onder de sterkst betrokkenen bij de school, dat wil zeggen onder ouders die lid zijn van de ouderraad of oudervereniging, bestuursleden, docenten die lid zijn van de MR en de schoolleiding. Per school werden drie ouders, drie docenten, drie bestuursleden en een lid van de schoolleiding verzocht een vragenlijst in te vullen. Per school leverde dat maximaal tien respondenten op, in totaal 170 voor het complete scholenbestand. Van deze 170 personen hebben 122 de vragenlijst geretourneerd.

Data-analyse

De drie bovengenoemde data-bronnen (interviews, document-analyse, outputlijst) zijn verwerkt tot een schoolspecifiek verslag. Dit verslag is ter correctie opgestuurd naar elke school. Vervolgens zijn de gecorrigeerde verslagen op kwalitatieve wijze geanalyseerd. Dit betekent dat vooral werd gezocht naar de gemeenschappelijke kenmerken van elk schooltype en de verschillen tussen de schooltypen. Om de conclusies zo goed mogelijk te kunnen verantwoorden is bij de analyses ook gebruik gemaakt van kwantificering. Vooral de outputlijst en de document-analyse (vooral van het PR-materiaal) hebben hier een rol gespeeld. Zo is voor de outputlijst het aantal activiteiten dat een school onderneemt per onderscheiden categorie geteld. Vervolgens is, voorzover vergelijkbaar, gekeken naar de verschillen tussen de verschillende schooltypen. Eenzelfde analyse is toegepast voor de analyse van de PR-materialen.

Om verschillen in houdingen tussen vreschillende groepen vast te stellen is gebruik gemaakt van een eenwegs-variantie-analyse. Als factoren dienden respectievelijk het schooltype waarbij de respondenten betrokken waren en de geleding waartoe zij behoorden. Als afhankelijke variabelen dienden de gemiddelde scores op de drie geconstrueerde schalen (deregulering, responsiviteit, kwaliteitsoordeel).

Resultaten

het beleid van scholen

Zoals bij de vraagstelling naar voren is gekomen, is voor de marktpositie van scholen van belang dat zij in staat moeten zijn een toekomstgerichte werkwijze te ontwikkelen tot uiting komend in een op de situatie van de school gericht onderwijs- en personeelsbeleid. Van belang daarbij zijn informatie over de omgeving, de leerlingmarkt, de kwaliteit van docenten en onderwijs en aandacht voor relevante doelgroepen als (potentiële) leerlingen, toeleverende en afnemende scholen en werkgevers. Het zijn deze punten die hieronder aan bod zullen komen.

Duidelijk is dat het moeilijk is gezien de kwalitatieve aard van de gegevens deze in tabelvorm weer te geven. Niettemin kan toch enig inzicht in de materie worden verkregen door tabel 3. Deze tabel is tot stand gekomen op basis van de output-vragenlijst en geeft met name een eerste inzicht in de veranderingen die scholen de afgelopen jaren op een aantal van de bovengenoemde terreinen hebben doorgevoerd.

Uit de resultaten uit Tabel 3 kan afgeleid worden dat binnen vrijwel alle schooltypen op bovengenoemde aspecten de afgelopen jaren de nodige veranderingen zijn opgetreden. In algemene zin kan gezegd worden dat scholen een meer responsieve houding ten opzichte van hun omge-

Tabel 3. Gemiddeld aantal activiteiten per deelgebied.

	BO		AVO		MBO	
	1987	nu	1987	nu	1987	nu
promotie						
voorlichting	1.17	1.33	6.00	5.83	6.00	5.80
reclame / p.r.	1.00	2.00	1.00	2.33	1.40	3.60
herkenbaarheid	1.33	3.50	4.00	5.17	5.40	8.40
externe contacten						
verbeteren contacten met ouders/leerlingen	3.17	3.67	5.00	5.50	2.60	3.20
vergroten medezeggenschap ouders/ leerlingen	1.50	2.33	2.00	2.50	2.20	3.00
contact toeleverende scholen	–	–	3.83	9.00	5.00	16.20
contact vervolgonderwijs	3.17	4.50	4.17	5.33	3.60	5.40
contact bedrijfsleven	–	–	–	–	3.60	5.40
beleid						
personeel	2.33	2.30	4.83	5.50	6.60	8.00
systematiek gegevens	2.00	2.83	3.67	4.67	2.80	4.60
beleidsplannen	1.50	3.33	1.17	3.83	1.40	4.40
onderwijs						
onderwijsaanpassing	2.00	4.67	2.83	4.17	4.40	5.80
andere veranderingen m.b.t. onderwijs	1.67	2.33	3.17	3.83	2.80	3.80

ving innemen, meer informatie over hun omgeving tot hun beschikking hebben en meer beleidsmatig te werk gaan.

Relevant in het kader van de responsiviteit zijn vooral de toegenomen contacten van scholen met doelgroepen als ouders, bedrijfsleven en toeleverende scholen. Vooral de toename in contacten van MBO-scholen met toeleverende scholen springt in het oog. Specifiek moet hier niet alleen gedacht worden aan het intensiveren van contacten met leerkrachten en schoolhoofden, maar vooral aan de toename van het fenomeen terugrapportage aan toeleverende scholen over de prestatie van ex-leerlingen.

Dit laatste wijst er al op dat MBO-scholen ook meer dan voorheen de beschikking hebben over relevante informatie. In de eerste plaats heeft deze betrekking op gegevens omtrent leerlingen zoals gegevens over de leerprestaties, rendement en vakken- en richtingkeuze binnen de school. In de tweede plaats weet men in toenemende mate ook waar leerlingen na het volgen van de opleiding terechtkomen. Dergelijke tendensen zijn ook bij de andere schooltypen te traceren (zie tabel 3 onder het kopje systematische gegevens). Daarnaast hebben vrijwel alle scholen in het onderzoek de beschikking over gegevens inzake andere relevante omgevingsontwikkelingen. Zo hebben vrijwel alle scholen prognoses over de te verwachten leerlingaantallen en de gebieden/buurten van afkomst tot hun beschikking. Verder zijn er bij de scholen in het middelbaar beroepsonderwijs aanzetten tot bredere omgevingsanalyses naspeurbaar waarbij systematisch geïnventariseerd wordt wat bijvoorbeeld decanen, werkgevers en/of ex-leerlingen van de opleiding vinden. Als zodanig zijn de onderzochte scholen dus in toenemende mate in staat op basis van deze kennis hun marktpositie te verstevigen en hun beleid er op af te stemmen.

Feitelijk kan hier gewezen worden op de toename van het aantal relevante beleidsplannen en van het aantal personen binnen de school dat specifiek belast is met bepaalde taken in het kader van het versterken van de marktpositie. Specifiek gaat het bij het laatste bijvoorbeeld om het institutionaliseren van contacten met het bedrijfsleven, het onderhouden van contacten met ouders, potentiële leerlingen, toeleverende scholen en de pers.

Een laatste tendens is tenslotte dat met name de scholen in het AVO en het MBO hun marktpositie proberen te verstevigen door veranderingen in te voeren in hun onderwijs. Bij scholen in het AVO worden daarbij vooral wegen gezocht om qua aanpak aan te sluiten bij een beoogde doelgroep, terwijl in het MBO veelvuldig gezocht wordt naar de mogelijkheid om nieuwe vakken/richtingen aan te bieden. Niet alleen om het onderwijs in toenemende mate af te stemmen op het bedrijfsleven, maar ook om het keuzeaanbod voor leerlingen zo groot mogelijk te maken met als doel het potentieel aan geïnteresseerde leerlingen te verhogen.

Tenslotte, geen eenduidige verschillen zijn er wat betreft het huidige functioneren op bovengenoemde aspecten gevonden tussen scholen met dalende dan wel stijgende leerlingaantallen. Wel kan geconcludeerd worden dat scholen met dalende leerlingaantallen hun activiteiten om hun marktpositie te verbeteren de afgelopen jaren meer geïntensiveerd hebben dan scholen met stijgende leerlingaantallen. Dit blijkt met name uit de resultaten op de output-lijst; de verschilscore tussen het aantal activiteiten die scholen in 1987 ondernamen en die ze nu ondernemen is steeds groter bij scholen in een bedreigde positie dan bij scholen met een constante of stijgende leerlinginstroom.

houdingen

De resultaten met betrekking tot de houdingen zijn wel eenduidig in tabelvorm weer te geven. In tabel 4 staan de gemiddelde scores van de verschillende geledingen op de drie ontwikkelde schalen vermeld.

Uit deze scores blijkt dat met name ouders een responsieve houding van de school het meest waarderen en docenten het minst. Dit verschil wordt vooral bepaald door de scores op de items die betrekking hebben op de betrokkenheid van ouders bij het onderwijs. Met name ouders zijn van mening dat scholen op dit punt beter moeten functioneren, waarbij zij vaak gesteund worden door bestuursleden. Specifiek kan hier gedacht worden aan het presenteren van jaarcijfers zodat ouders beter tussen scholen kunnen kiezen en een grotere betrokkenheid van ouders bij het onderwijs. Met name docenten denken hier vaak minder positief over. Een dergelijk verschil

Tabel 4. Houdingen van verschillende betrokkenen ten opzichte van het streven naar meer responsiviteit, deregulering en de verwachte gevolgen van deregulering voor het functioneren van scholen.

	ouders (n=30)	school-leiding (n=18)	docenten (n=41)	bestuur (n=31)	F-ratio	p-waarde
responsiviteit	4.20	3.91	3.76	4.06	4.23	0.00
deregulering	2.44	2.86	2.56	2.90	3.85	0.01
kwaliteitsoordeel schaalrange 1-5	2.97	3.76	3.17	3.64	4.88	0.00

doet zich ook voor als gekeken wordt naar de verhouding schoolleiding – docenten. Zo zien docenten wel de noodzaak en wenselijkheid van profilering in, maar willen niet dat dit ten koste gaat van hun autonomie; een grotere invloed van schoolleiding op het onderwijs en het naschoolingsbeleid -als middel om te komen tot een meer duidelijke profilering van de school-, wordt door weinig docenten als positief ervaren. Op de dereguleringsschaal zijn ook verschillen tussen de verschillende geledingen te onderkennen. Hier zijn ouders de meest 'conservatieve' groep en voorstander van een sterke regulering, terwijl met name de bestuursleden daar iets soepeler over denken. Zonder overigens overtuigend voorstander van dereguleringsmaatregelen te zijn.

Hetzelfde geldt min of meer voor de ingeschatte gevolgen (kwaliteitsoordeel). Ouders denken negatiever over de gevolgen dan de schoolleiders en bestuursleden. Deze laatste twee groepen zijn zelfs opmerkelijk positief over de verwachte gevolgen; door het stimuleren van marktgerichtheid verwachten zij kwaliteitsverbetering van het onderwijs.
Uit Tabel 5 kunnen de verschillen in houdingen tussen betrokkenen bij de verschillende schooltypen afgelezen worden.

Uit deze tabel blijkt dat de verschillen tussen schooltypen met name zitten in de mate waarin deregulering gewenst wordt. Binnen het basisonderwijs en AVO zijn weinig voorstanders van deregulering, terwijl deregulering meer gewaardeerd wordt binnen het MBO. Deze verschillen worden vooral verklaard door aspecten als de mening over contractonderwijs, sponsoring, mogelijkheden tot een flexibel personeelsbeleid (met name differentiatie in salaris tussen goede en/ of schaarse docenten en anderen binnen de school) en de vrijheid om meer dan nu het geval is het onderwijsaanbod te bepalen. Het zijn met name betrokkenen bij het MBO die hier voorstander van zijn, terwijl betrokkenen bij het basisonderwijs hier de nodige bezwaren tegen koesteren.

Tabel 5. Houdingen binnen verschillende schooltypen ten opzichte van het streven naar responsiviteit, deregulering en de verwachte gevolgen voor het functioneren van scholen.

	Basisschool (n=49)	AVO (n=35)	MBO (n=38)	F-ratio	p-waarde
responsiviteit	3.88	4.03	4.01	0.85	0.42
deregulering	2.35	2.55	3.16	2.5	0.00
kwaliteitsoordeel schaalrange 1-5	3.19	3.20	3.59	2.47	0.00

HET TWEEDE ONDERZOEK: RENDEMENT

Inleiding

De overheid overweegt in het kader van het dereguleringsbeleid invoering van een vorm van rendementsbekostiging in de bovenbouw van het HAVO- en VWO als middel om scholen te stimuleren hun rendement te verbeteren. Met een financiering op basis van rendement verplaatst de aandacht van de overheid zich per definitie van de input en procesbewaking naar de outputkant van het onderwijs. Voordat deze overheidsplannen verder uitgewerkt worden is eerst het draagvlak voor deze maatregel in het onderwijsveld gepolst en zijn de mogelijke beleidsreacties van scholen geïnventariseerd. Dit onderzoek naar het oordeel van de schoolleiders over rendementsbekostiging wordt hieronder kort beschreven.

Selectie van scholen en respons

Voor de selectie van scholen is gebruik gemaakt van de integrale leerlingtellingen. Alle HAVO- en VWO-scholen zijn ingedeeld op de kenmerken denominatie (openbaar of bijzonder), schoolsoort (HAVO of VWO), urbanisatiegraad (≤100.000 inwoners of >100.000 inwoners), schooltype (HAVO/VWO-scholengemeenschap of anderszins samengestelde scholengemeenschap) en rendement (laag, middelmatig of hoog rendement). Het rendement is berekend door per school de ratio van percentage bevorderden in de bovenbouw plus het percentage geslaagden op het toaal aantal leerlingen in de bovenbouw te berekenen. Het rendementscriterium is van belang omdat verwacht kan worden dat de reactie van scholen op rendementsbekostiging afhankelijk is van het eigen, huidige rendement. De combinatie van de vijf selectiecriteria resulteert in een tabel met 48 cellen (2*2*2*2*3). Er is gestreefd naar een bezetting van één school per cel. In eerste instantie werd een school benaderd. Bij weigering tot deelname werd een nieuwe school benaderd. Naast deze scholen zijn acht scholen met een hoge concentratie culturele minderheden benaderd. Het bleek niet mogelijk voor iedere combinatie een school te vinden. Zo komen bijzondere scholen met een zeer laag rendement nauwelijks voor. In totaal zijn 94 scholen benaderd, waarvan er 53 hun medewerking hebben verleend. Dit betekent een responspercentage van 56%. Hoewel de selectie niet representatief kan worden genoemd voor de populatie HAVO- en VWO-scholen in Nederland, heeft de gekozen gestratificeerde benadering het voordeel dat zinvol vergelijkingen kunnen worden gemaakt tussen sub-groepen van scholen.

Onderzoeksprocedure

De gegevens zijn verzameld met behulp van een schriftelijke vragenlijst en een interview. De schriftelijke vragenlijst diende voor het verzamelen van gegevens over leerlingaantallen, slaag- en doublurepercentages en over het huidige beleid van scholen inzake rendementsbewaking. Gegevens over het huidige beleid zijn verzameld om een vergelijking te kunnen maken met de voorgenomen maatregelen indien een vorm van rendementsbekostiging zou worden ingevoerd. In de interviews met de conrectoren is ingegaan op het oordeel van over rendementsbekostiging, alsmede op beleidsmaatregelen die de school zou treffen bij invoering van rendementsbekostiging. Om de respondent in de gelegenheid te stellen intern overleg te voeren over de mogelijke reacties van de school op invoering van rendementsbekostiging zijn de interviewonderwerpen vooraf aan de respondent meegedeeld. De respondenten zijn geconfronteerd met de financiële consequenties voor de eigen school van invoering van vier optionele bekostigingsmodellen, met de vraag daaraan gekoppeld wat zij zouden doen bij invoering van één van deze modellen. Een probleem bij deze wijze van werken is het gevaar van sociale wenselijkheid. Met andere woorden, durven respondenten ongewenste maatregelen te berde te brengen. Om dit te voorkomen zijn een aantal maatregelen genomen. In de eerste plaats is een glijdene schaal geconstrueerd (zie volgende paragraaf) met daarin opgenomen vele gradaties van niet wenselijke antwoorden. Daarnaast is de vraagstelling aangepast door niet te vragen of de school de maatregel zou nemen, maar hoe groot men de kans op invoering van een dergelijke maatregel inschat. Vervolgens is gebruikt gemaakt van 'dubbele bevraging'. Aan respondenten is eerst op algemeen niveau en vervolgens op concreet niveau dezelfde vraag gesteld. Hun antwoorden bleken con-

sistent. Hoewel hiermee de kans op sociale wenselijkheid niet volledig weggenomen is, mag op basis van de literatuur (Swanborn, 1987) wel verwacht worden dat de kans hierop sterk verminderd is.

Tenslotte is de kans op sociaal wenselijke antwoorden ondervangen door de helft van de respondenten eerst met het meest ingrijpende bekostigingsmodel (50%-model) te confronteren en vervolgens met de minder ingrijpende modellen. De andere helft respondenten werd eerst met het minst ingrijpende bekostigingsmodel (5%-model) geconfronteerd en vervolgens met de meer ingrijpende. De antwoordpatronen van beide groepen schoolleiders bleken niet te verschillen.

Instrumentontwikkeling en data-analyse

Het voorgenomen beleid van scholen is getypeerd met de termen strategisch beleid en strategisch gedrag. *Strategisch beleid* betreft de beleidsvoering die enerzijds gericht is op de verbetering van het rendement van de eigen school en anderzijds bijdraagt aan de realisatie van de overheidsdoelen, namelijk het in samenhang vergroten van het kwalitatief en het kwantitatief rendement van de bovenbouw van het voortgezet onderwijs. Een belangrijk criterium is dat deze maatregelen zodanig ten goede komen aan de leerlingen dat uiteindelijk het niveau van de leerprestaties wordt verhoogd. Strategisch beleid verwijst aldus naar maatregelen die leiden tot de gewenste effecten voor het voortgezet onderwijs. *Strategisch gedrag* kan daarentegen meer getypeerd worden als ontwijkingsgedrag, gericht op budgetmaximilisatie, waarbij het belang van de leerling niet voorop staat. Het betreft de beleidsvoering die weliswaar gericht is op rendementsverbetering van de eigen school, doch niet bijdraagt aan de realisatie van de overheidsdoelen ten aanzien van de verbetering van zowel het kwantitatieve als het kwalitatieve rendement van de bovenbouw van het voortgezet onderwijs. Strategisch gedrag verwijst aldus naar beleidsmaatregelen die leiden tot ongewenste neveneffecten. We kunnen hier te maken hebben met ad hoc beslissingen als reactie op een wellicht tegenvallende bekostiging. Voorbeelden hiervan zijn ingrepen in het schoolonderzoek (verlagen van niveau door vereenvoudiging van opgaven, het laten leren voor de toets of het achteráf aanpassen van de norm) of een tijdelijke verscherping van de doorstroming van de onderbouw naar de bovenbouw (bijstellen overgangsnormen, 'bespreekzone' voor twijfelgevallen versmallen of verzwaren proefwerken). Het kan ook een weloverwogen lange-termijnbeleid, gericht op regulatie van de instroom betreffen. Om de mogelijke beleidsreacties op één van de optionele bekostigingsmodellen te kunnen weergeven in termen van strategisch beleid en strategisch gedrag zijn twee schalen geconstrueerd. De constructie van de eerste schaal leverde de volgende statistische gegevens op (Tabel 6).

Tabel 6. Ontvouwingsgegevens.

		H(I)	T(H)
strategisch beleid	▲		
	d – optimaliseren functioneren organisatie	.62	3.04
	a – optimalisaren leerlingbegeleiding	.41	4.21
	e – stimuleren (vak)inhoud. verbeteringen	.42	5.34
	g – veranderen schoolwerkplan	.40	5.51
	c – optimal. selectie- en determinatiebeleid	.31	4.29
	j – verscherpen aannamebeleid	.34	3.73
	i – verscherpen doublurebeleid	.39	3.20
	h – verscherpen doorstroombeleid	.22	1.39
strategisch gedrag	▼		
Totale schaal		.38	4.79

De H(I)-coefficient geeft aan in hoeverre een maatregel schaalbaar is. De coëfficiënt loopt van 0 (niet schaalbaar) tot 1 (perfect schaalbaar). In het algemeen duidt een waarde van .30 op voldoende en .40 op redelijke schaalbaarheid. Daarnaast is de T-waarde vermeld, die wanneer groter dan 1.96, aangeeft dat de geschatte H(I)-waarde statistisch van 0 (niet schaalbaar) afwijkt met 5% kans op een verkeerde beslissing (in werkelijkheid was de maatregel niet schaalbaar). De totale schaal is voldoende tot redelijk homogeen; het verwijderen van item h (verscherpen doorstroombeleid) leidt niet tot een verbetering, zodat dit item ondanks de matige itemkarakteristiek gehandhaafd blijft.

Naast het aangeven van beleidsprioriteiten is scholen gevraagd naar concrete maatregelen die genomen zouden worden bij invoering van rendementsbekostiging. Hiervoor is een Likertschaal ontwikkeld voor de dimensie van strategisch beleid tot strategisch gedrag. De betrouwbaarheid van deze schaal is goed (Cronbach's $\alpha = 0{,}78$). De verzamelde data voor het totale onderzoek zijn geanalyseerd met behulp van rechte tellingen, variantie-analyses, correlatietechnieken en schaaltechnieken.

Resultaten

huidig beleid inzake bewaking van het rendement

Scholen verschillen in hun huidige beleidsvoering ten aanzien van de aanname, de selectie en determinatie en de leerlingbegeleiding en deze verschillen hangen samen met hun rendement. HAVO-scholen met een hoog rendement baseren de aanname van de leerlingen in de onderbouw en in de bovenbouw meer op gebleken prestaties dan op criteria als voorkeur van leerlingen en/of ouders. HAVO-scholen met een laag rendement blijken bij de advisering over de vakkenpakketkeus risico's te vermijden; er wordt liever te laag dan te hoog geadviseerd. Bij VWO-scholen treden deze verbanden niet op. HAVO-scholen met een hoog rendement hebben verhoudingsgewijs minder verschillende voorzieningen voor leerlingbegeleiding, maar besteden gemiddeld wel meer uren aan die begeleiding per leerling. VWO-scholen met een hoog rendement blijken juist meer verschillende voorzieningen voor leerlingbegeleiding te hebben terwijl er geen significante verschillen qua tijdsbesteding optreden.

oordeel

Er zijn geen verschillen aangetroffen tussen scholen in hun houdingen ten opzichte van rendementsbekostiging. Scholen oordelen over het algemeen tamelijk negatief over de invoering van een dergelijk bekostigingsstelsel, hetgeen geïllustreerd wordt in Tabel 7.

Respectievelijk 64.2% en 60.4% van de scholen is tegen rendementsbekostiging in het algemeen, dan wel tegen rendementsbekostiging voor alleen de bovenbouw. Er worden met name negatieve gevolgen voor de leerlingen gevreesd. Zwakke leerlingen en laatbloeiers zullen minder kansen krijgen een diploma te halen, zo oordelen de schoolleiders. Het minst negatief wordt

Tabel 7. Oordeel over rendementsbekostiging.

	negatief %	neutraal %	positief %	N
oordeel over rendements-bekostiging in het algemeen	64.2	20.8	15.0	53
oordeel over rendements-bekostiging voor de bovenbouw	60.4	24.5	15.1	53
oordeel over rendements-bekostiging met bonus voor probleemgroepen	46.5	25.8	27.6	48

geoordeeld over een vorm van rendementsbekostiging waarbij een bonus voor geslaagden uit achterstandsgroepen, zoals culturele minderheden, wordt uitgekeerd (ruim 27% oordeelt positief en circa 26% heeft een neutraal oordeel). Het begeleiden van leerlingen in achterstandssituaties vraagt een extra prestatie van scholen en deze mag beloond worden, zo vindt men.

Schoolleiders gaven ook hun oordeel over mogelijke andere consequenties voor de school van invoering van rendementsbekostiging. Met name het pedagogisch/didactisch klimaat en de omgang met leerlingen zullen volgens hen leiden onder deze maatregel. Daarnaast komen volgens de helft van de schoolleiders de vele extra-curriculaire activiteiten, zoals meerdaagse excursies en creatieve activiteiten onder druk te staan. Positieve consequenties zou de maatregel kunnen hebben voor de professionaliteit van de organisatie, de efficiëntie van het werk en de marketing en de Public Relations.

Om enig idee te krijgen van het rendementsbewustzijn van de schoolleider is hierover een aantal vragen gesteld. Op de vraag of men de slaagpercentages wel eens vergelijkt, antwoordt 96% met 'ja' en 4% met 'nee' (N=47). Op de vraag of men het landelijk gemiddelde slaagpercentage 1989/1990 kent antwoordt 75% van de HAVO-scholen (N=24) en 87% van de VWO-scholen (N=23) bevestigend. Vervolgens is aan de schoolleiders eerst het rendementscijfer van de school gepresenteerd en daarna gevraagd aan te geven hoe men denkt dat het rendement van de school zich tot het landelijk gemiddelde verhoudt. De antwoorden van de respondenten geven aan dat het rendementsbewustzijn niet erg groot is; slechts 50% van de betrokkenen heeft een juist inzicht van het rendement van de school in verhouding tot het landelijk gemiddelde. Er is verder gekeken of dit rendementsbewustzijn samenhangt met de aard van het oordeel over rendementsbekostiging. Dit blijkt niet het geval te zijn.

voorgenomen beleidsmaatregelen

De beleidsreacties van scholen zijn onderscheiden naar algemene en concrete maatregelen en kunnen worden afgebeeld op de dimensie van strategisch gedrag naar strategisch beleid, respectievelijk met behulp van de ontvouwingschaal en de Likertschaal. De schaalscores voor de algemene beleidsmaatregelen lopen van 0 (strategisch gedrag) tot 7 (strategisch beleid). De schaalscores voor de concrete maatregelen lopen van 1 (strategisch gedrag) tot 5 (strategisch beleid). Er zijn variantie-analyses uitgevoerd van de scores op de ontvouwingsschaal (algemeen beleid) en van de scores op de Likertschaal (concrete maatregelen) als afhankelijke variabelen met de factoren denominatie (openbaar of bijzonder), urbanisatiegraad (<100.000 inwoners of >100.000 inwoners)), rendement (laag, middelmatig of hoog rendement), schooltype (HAVO/VWO-scholengemeenschap of anderszins samengestelde scholengemeenschap), onderwijssoort (HAVO of VWOs) en oordeel over rendementsbekostiging (negatief, neutraal of positief). Vier analyses leverden significante verbanden op. De gevonden significante verbanden staan vermeld in Tabel 8.

Scholen met een hoog rendement zijn relatief meer geneigd te kiezen voor algemene beleidsmaatregelen die tot het strategisch beleid kunnen worden gerekend, dan scholen met een laag rendement. Hetzelfde patroon blijkt uit de positie van de scholen op de Likertschaal, die is geconstrueerd op basis van de concrete maatregelen. Bijzondere scholen, scholen met een hoog rendement en de HAVO-scholen zijn iets meer geneigd tot strategisch beleid dan de openbare scholen, de scholen met een laag rendement en de VWO-scholen. Echter, scholen neigen over het algemeen meer naar strategisch beleid dan naar strategisch gedrag. Extreme vormen van strategisch gedrag, zoals versoepeling van de bewaking van het kwalitatief niveau (bijvoorbeeld verlagen van de moeilijkheidsgraad van de schoolonderzoeken), worden amper of niet als beleidsoptie genoemd. Het probleem van de sociale wenselijkheid bij het geven van antwoorden is waarschijnlijk, ondanks de aanpassingen in de vraagstelling, niet geheel opgelost. Vermeld dient echter te worden dat scholen beseften dat hun antwoorden van belang zijn voor de verdere beleidsvorming rond rendementsbekostiging. Gezien hun negatieve oordeel over deze vorm van outputcontrole zou men ook kunnen verwachten dat scholen juist extreme vormen van strategisch gedrag noemen, teneinde de voorgenomen maatregelen inzake rendementsbekostiging af

Tabel 8. Strategisch beleid en strategisch gedrag op ontvouwingsschaal (algemene maatregelen) en Likert-schaal (concrete maatregelen) naar denominatie, rendement en onderwijssoort.

	Algemene maatregelen		Concrete maatregelen		
	Gem.	SD	Gem.	SD	N
Denominatie [1]					
Bijzonder			3.17	0.47	23
Openbaar			2.97	0.41	30
Rendement [2]					
laag	3.82	1.33	2.90	0.43	17
midden	4.23	1.60	2.98	0.42	17
hoog	5.18	1.69	3.26	0.41	19
Onderwijssoort [3]					
HAVO			3.17	0.36	28
VWO			2.93	0.50	25

[1] $p = 0.10$; $Eta^2 = 0.05$
[2] $p = 0.03$; $Eta^2 = 0.13$
[3] $p = 0.04$; $Eta^2 = 0.08$

te wentelen. Een empirische indicatie voor het feit dat extreme vormen van strategisch gedrag niet direct te verwachten zijn, betreft de in eerder onderzoek (Bosker, 1991) aangetroffen (licht) positieve correlatie tussen kwalitatief en kwantitatief rendement. Hogere prestaties gaan in de praktijk samen met minder zittenblijven, minder drop-out en minder afstroom.

Wat betreft de concrete maatregelen, blijkt dat meer dan 50% van de scholen het zeer waarschijnlijk acht dat bij invoering van een vorm van rendementsbekostiging de volgende maatregelen zullen worden getroffen:

1. creëren van meer bijscholingsmogelijkheden voor docenten;
2. invoeren dan wel automatiseren van een leerlingvolgsysteem;
3. stoppen van meer tijd per leerling in studiebegeleiding;
4. voorzichtiger adviseren bij twijfel over schooltypekeuze;
5. vaker doorstroming van HAVO-3 naar het MBO adviseren;
6. sneller de afstroom naar een lager schooltype adviseren.

veranderingen in beleid

De maatregelen die de scholen zeggen te gaan nemen bij invoering van een vorm van rendementsbekostiging, zijn vergeleken met het huidige beleid dat de school voert, teneinde inzicht te krijgen in de veranderingen in beleid. Tabel 9 toont dat scholen meer geneigd zijn strategische beleidsmaatregelen ter verbetering van hun rendement te treffen naarmate het organisatie- en onderwijskundige beleid op de corresponderende terreinen in mindere mate ontwikkeld is. Dit komt meer specifiek tot uitdrukking bij aspecten van de leerlingbegeleiding, de leerlingbeoordeling en de schoolorganisatie. Hoe minder faciliteiten men in de onderbouw voor de leerlingbegeleiding kent, des te eerder kiest men voor beleidsterreinen en concrete maatregelen die kunnen worden gerekend tot strategisch beleid. Verder geldt dat scholen verhoudingsgewijs meer kiezen voor de strategische beleidsterreinen, naarmate ze in de huidige situatie minder maatregelen kennen om ongeoorloofd verzuim van leerlingen tegen te gaan en naarmate hun beoordelingssysteem minder systematisch is.

Tabel 9. Samenhang tussen aspecten van huidige en toekomstige schoolsituatie.

Huidig beleid	Voorgenomen algemene maatregelen	Voorgenomen concrete maatregelen
AANNAME		
toelatingscriteria onderbouw	-.10	*.16*
toelatingscriteria bovenbouw HAVO	*-.18*	-.01
toelatingscriteria bovenbouw VWO	-.17	*-.27*
LEERLINGBEGELEIDING		
voorzieningen onderbouw	**-.27**	**-.23**
voorzieningen bovenbouw	-.09	-.01
aantal uren onderbouw	.02	*.16*
aantal uren bovenbouw	-.02	.12
LEERLINGBEOORDELING		
beoordelingssysteem	**-.25**	-.08
DOORSTROMING		
criteria advies na brugklas	*-.13*	*-.15*
SCHOOLORGANISATIE		
aantal functioneringsgesprekken	-.03	-.06
verzuimmaatregelen	**-.33**	*-.15*
SELECTIE/DETERMINATIE		
advisering vakkenpakketkeuze	.01	.13
BESTUURLIJKE BETROKKENHEID:		
organisatiebeleid	.01	.11
onderwijskundig beleid	*-.13*	*-.13*
PRESTATIEGERICHTHEID	-.10	.02

vet gedrukte correlaties zijn significant bij $\alpha < .10$
cursief gedrukte correlaties zijn significant bij $\alpha < .20$

CONCLUSIES

In dit artikel staan twee vragen centraal. De eerste vraag luidt of scholen komen tot een marktgericht beleid zoals bedoeld door het dereguleringsbeleid en welke problemen zij daarbij ondervinden. De tweede vraag heeft betrekking op de houdingen aangaande de wenselijkheid van marktgericht functioneren van scholen en een overheidsbeleid gericht op het stimuleren hiervan. Deze vragen zullen achtereenvolgens besproken worden.

Vanuit het eerste onderzoek gesteld kan geconcludeerd worden dat scholen door het dereguleringsbeleid inderdaad gestimuleerd worden beleidsmatiger te werken. Dit gegeven kan aan de hand van een aantal ontwikkelingen nader geïllustreerd worden. In het basisonderwijs heeft het formatiebudgetsysteem er volgens de betrokkenen toe geleid dat scholen steeds meer beschikken over prognoses over de verwachte leerlinginstroom. In de onderzochte VWO-scholen is men druk bezig met het vormgeven van de vrije ruimte in het onderwijsaanbod. In de onderzochte MBO-scholen heeft volgens de betrokkenen de lumpsum-systematiek geleid tot het creëren van nieuwe opleidingen, flexibilisering in de doorstroommogelijkeden en het creëren van nieuwe functies dan wel het differentiëren van functies. Een voorbeeld hiervan is dat alle bezochte MBO-scholen in meer of mindere mate taakuren hebben vrijgemaakt om een beter

beleid aangaande aspecten die van groot belang zijn voor de marktpositie (bijvoorbeeld leerlingwerving) mogelijk te maken. Tenslotte valt bij alle schooltypen te wijzen op de toename van het aantal contacten met specifieke doelgroepen.

Deze conclusie wil niet zeggen dat scholen geen problemen ondervinden. In algemene zin kan gesteld worden dat het proces van deregulering en decentralisatie nog in een overgangsfase verkeert. Dit levert volgens de betrokkenen een aantal problemen op. In de eerste plaats is er de spanning tussen het stimuleren van de marktgerichtheid van scholen aan de ene kant en aan de andere kant de belemmeringen die scholen ondervinden om aan deze marktgerichtheid handen en voeten te geven. Een voorbeeld hiervan is dat de MBO-scholen in dit onderzoek willen concurreren met betrekking tot het onderwijsaanbod, maar daarbij in de praktijk allerlei belemmeringen ondervinden, met name omdat bij de verdeling van vakken/richtingen over scholen vanuit de overheid overwegingen ten aanzien van de denominatie een belangrijke rol spelen. Een ander problematisch punt in deze is dat, ondanks het formatiebudgetsysteem, de mogelijkheden om te komen tot een flexibele personeelsbeleid nog altijd als gering ingeschat worden. Met name de verplichting om bij vacatures wachtgelders in dienst te nemen is hier de boosdoener.

In de tweede plaats vloeien er knelpunten voort uit het feit dat er in veel scholen sprake is van een cultuur die niet toegesneden is op de veranderende omstandigheden. Uit de interviews met de schoolleiders en uit de vragenlijst blijkt dat de houding van docenten in veel gevallen niet in de pas loopt met een marktgerichte filosofie. Niet alleen zijn zij weinig geneigd om deel te nemen aan de beleidsvoering, maar tevens houden zij veelal vast aan hun opvattingen aangaande autonomie. Duidelijk is dat in een dergelijke constellatie de vorming en uitvoering van instellingsbeleid niet eenvoudig is.

De laatste conclusies hebben betrekking op de houdingen binnen de onderzochte scholen. Met betrekking tot de responsiviteit binnen scholen kan geconcludeerd worden dat deze in het algemeen groter is dan de theorievorming op dit terrein doet vermoeden. Zo levert een sterke gerichtheid van scholen in het beroepsonderwijs op het bedrijfsleven voor de meeste respondenten geen problemen op. In dit opzicht zijn opvattingen die vasthouden aan de 'pedagogische provincie' blijkbaar aan het verdwijnen. Problematisch is wel de verhouding tot ouders. Niet alleen schatten betrokkenen de deskundigheid van ouders met betrekking tot het onderwijs laag in, maar tevens wijzen veel docenten een grote betrokkenheid van ouders bij het onderwijs af. Voorzover het dereguleringsbeleid als doel heeft de afstemming tussen het onderwijsaanbod van scholen en de wensen van onderwijsconsumenten te vergroten, moeten de kansen hierop als klein ingeschat worden.

Wat betreft de houdingen aangaande het overheidsbeleid kan geconcludeerd worden dat er tussen de verschillende schooltypen grote verschillen bestaan over de wenselijkheid van deregulering. In het basisonderwijs is men in grote lijnen geen sterk voorstander hiervan, in het MBO wel. Dit zal betekenen dat maatregelen in het kader van het dereguleringsbeleid in de verschillende schooltypen verschillend zullen uitwerken. Een voorbeeld hiervan is de mogelijkheid om te differentiëren naar salaris. In het basisonderwijs is hier weinig behoefte aan, in het MBO veel meer. Dat dit consequenties heeft voor de uitwerking van dit beleid in de verschillende schooltypen ligt voor de hand. Getuige de grote verschillen tussen schooltypen ligt het evenzeer in de rede om bij een nadere uitwerking van het dereguleringsbeleid met deze verschillen in oriëntatie rekening te houden.

Specifieke inzichten omtrent de houdingen kunnen ontleend worden aan het onderzoek naar de wenselijkheid van rendementsbekostiging. Een belangrijke conclusie uit dit onderzoek is dat scholen afwijzend staan tegenover de invoering van een dergelijk systeem. Met name de notie dat vooral kenmerken van de leerlinginstroom bepalend zijn voor het rendement van de school speelt hier een belangrijke rol. Ondanks dit negatieve oordeel over de voorgestelde vormen van rendementsbekostiging hebben scholen echter wel vertrouwen in de mogelijkheden om hun rendement te manipuleren. Uit de genoemde te verwachten beleidsreacties van HAVO- en VWO scholen blijkt dat scholen gevoelig zijn voor bekostigingsprikkels en bij de invoering van een dergelijk systeem hun gedrag zullen aanpassen zonder te vervallen in extreme vormen van strategisch gedrag. Zoals gezegd kan sociale wenselijkheid hier een rol spelen. Daar staat echter

tegenover dat scholen in het MBO als reactie op rendementsbekostiging inderdaad wenselijk geachte maatregelen nemen gericht op optimalisering van de leerlingbegeleiding, -beoordeling en -selectie (Brandsma & Thijssen, 1994; Pelkmans & De Vries, 1991). Aangezien er geen reden is om aan te nemen dat de onderzochte scholen op dit punt wezenlijk van MBO-scholen verschillen, mag aangenomen worden dat scholen inderdaad conform de boven weergegeven reactie zullen handelen.

Een probleem inzake het verbinden van financiële consequenties aan een systeem van outputcontrole is echter wel dat het een technisch complexe maatregel is. Te wijzen valt hier op de noodzaak om met uitermate verfijnde scenario's te werken en op het bestaan van vele valkuilen die bij een gebrekkige imlementatie kunnen optreden. In dit opzicht verdient het aanbeveling om eerst een wat terughoudender middel in te zetten. Te denken valt hier aan de verplichting tot (openbare) jaarverslagen. Getuige het eerste onderzoek bestaan tegen een dergelijke maatregel in het onderwijs veel minder bezwaren. Bovendien wordt zo aan een sterk bij ouders levende wens tegemoet gekomen worden.

NOOT

1. Een nodaal is een gebied waarvan de bewoners in het maatschappelijk verkeer, waaronder de deelneming van onderwijs, geörienteerd zijn op bepaalde kernen in dit gebied. Een onderwijskern is een gemeente waar alle vormen van onderwijs die direct aansluiten op het lager onderwijs, aanwezig zijn (C.B.S., 1990).

LITERATUUR

Amelsvoort, H.W.C.H. van, Bosker, R.J., & Brandsma, T.F. (1992). *Outputfinanciering: pro's en contra's.* Enschede: Universiteit Twente, OCTO.

Berg, R. van den (1992). Baas in eigen school; innovatie als een strategisch vraagstuk binnen school. *Meso, 61*, 2-5.

Boef-van der Meulen, S., Bronneman-Helmers, R. & Konings-van der Snoek, M. (1983). *Schoolkeuzemotieven en meningen over onderwijs.* Rijswijk: Sociaal en Culutreel Planbureau.

Boyd, W.L. & Crowson, W.L. (1985). The changing conception and practice of public school school administration. *Review of Research in Education, 9*, 311-373.

Bosker, R.J., Guldemond, H., Hofman, R.H., & Hofman, W.H.A. (1988). *Kwaliteit in het voortgezet onderwijs.* Groningen: RION.

Bosker, R.J. (1991). De keuze van een onderwijsindicator; criterium, meetniveau en statistisch model. *Tijdschrift voor Onderwijswetenschappen, 21*, 191-205.

Brandsma, T.F. & Thijssen, J.A. (1994). *Flexibiliteit in de kwalificatievoorziening. Een onderzoek naar organisatorische en onderwijskundige kenmerken van een flexibiliseringsstrategie in het kader van de herstructurering van het beroepsonderwijs.* Enschede: OCTO (in voorbereiding).

C.B.S. (1989). *Loop van de bevolking per gemeente.* 's Gravenhage: SDU/CBS publicaties.

C.B.S. (1990). *Het voortgezet onderwijs regionaal bezien ('88-'89).* 's Gravenhage: SDU/CBS publicaties.

Chubb, J.E. & Moe, T.M. (1990). *Politics, markets and American schools.* Washington D.C.: Brookings Institution.

Commissie Evaluatie Basisonderwijs (1994). *Zicht op kwaliteit.*'s Gravenhage: SDU.

Dulmers, R.J. (1987). Het belang van onderwijsmarketing. In R.J. Dulmers & C. Dietvorst. *Marketing voor scholen.* Alphen aan de Rijn/Brussel: Samson Uitgeverij.

Faes, W. & Tilborgh, C. van (1983). *Marketing van diensten, een handboek voor praktisch management.* Deventer: Kluwer.

Gennip, J. van (1991). *Veranderingscapaciteiten van basisscholen.* Nijmegen: ITS.

Geurts, J. (1989). *Van Niemandsland naar beroepenstruktuur. Een studie over de aansluiting tussen onderwijs en arbeid op het niveau van aankomend vakmanschap* (proefschrift), Nijmegen: ITS.

Gielen, K.J.M. (1987). Autonomie van scholen en schoolmanagement-ontwikkeling. In L. Genemans, (ed.). *Autonomie van scholen en deregulering: verslag van een studieconferentie, Nijmegen, 6 februari 1987* (pp. 99-107) Nijmegen: ITS.

Giesbers, J.H.G.I. (1987). Autonomie en deregulering; ook een kwestie van kunnen en/of willen. In L. Genemans, (ed.). *Autonomie van scholen en deregulering: verslag van een studieconferentie, Nijmegen, 6 februari 1987.* (pp. 153-169) Nijmegen: ITS.

Giesbers, J.H.G.I. (1990). Autonomievergroting: lust of last? In J.H.G.I. Giesbers & P.M.E.F. Adriaens [et al.]. *De noodzaak tot verantwoording: beschouwingen over autonomie en evaluatie in het onderwijs.* (pp. 8-20) Algemeen Bureau Katholiek Onderwijs, Thema 1. Den Haag.

Graaf, H. v.d. & Hoppe, R. (1991). *Beleid en politiek.* Muiderberg: Coutinho.

Hoof, J.J. van (1986). *Aansluiting tussen onderwijs en arbeidsmarkt: recente ontwikkelingen en uitgangspunten voor beleid.* 's Gravenhage: Organisatie voor Strategisch Arbeidsmarktonderzoek.

Jakobs, J.J. (1987) Opvattingen in het onderwijsveld over grotere autonomie van scholen. In L. Genemans (ed.). *Autonomie van scholen en deregulering: verslag van een studieconferentie, Nijmegen, 6 februari 1987.* (pp. 81-91) Nijmegen: ITS.

Kotler, P. & Fox, K.F.A. (1985). *Strategic marketing for educational institutions.* Englewood Cliff/ New Jersey: Prentice Hall.

Leijenhorst, H. van (1993). *Kamerstuk 19790, nr 55.* (Verslag van een mondeling overleg). 's Gravenhage: SDU.

Lieberman, M. (1989). *Privatization and Educational Choice.* New York: St. Martin's Press.

Mahieu, P. (1985). *Schoolbank te koop: over schoolreclame en onderwijsmarketing.* Leuven:Acco.

Marx, E.C.H. (1987). Vermogen van scholen tot het voeren van bestuurlijk beleid. In L. Genemans (ed.). *Autonomie van scholen en deregulering: verslag van een studieconferentie, Nijmegen, 6 februari 1987.* (pp. 7-27) Nijmegen: ITS.

Ministerie van Onderwijs en Wetenschappen (1988). *De school op weg naar 2000.* Zoetermeer: Ministerie van Onderwijs en Wetenschappen.

Mintzberg, H. (1979). *The structuring of organizations.* Englewood Cliffs: Prentice Hall.

Pelkmans, A.H.W.M., van Kessel, N., van Viersen, D., Hahn, A., & Buurke, W. (1983). *Wensen omtrent scholen en de onderwijsplanning; vijf deelonderzoekingen over kwalitatieve onderwijsbehoeften, 'verlangd onderwijs' en de planning van het voortgezet onderwijs.* Nijmegen: ITS.

Pelkmans, A.H.W.M. & Vries, B. de (1991). *Modulering/flexibilisering in de tweedefase SVM.* Nijmegen: ITS. Voordracht gehouden op de conferentie 'Flexibiliteit in de kwalificatievoorziening', in het kader van het netwerk Beleidsonderzoekers Beroepsonderwijs en Volwasseneducatie, Utrecht, 24 oktober 1991.

Pieters, L. (1992). Trends in schoolkeuze na de basisschool. *Meso, 64,* 25 -30.

Ritzen, J.M.M. (1993). *Kamerstuk 19790, nr 54.* (Brief van de Minister van Onderwijs en Wetenschappen). 's Gravenhage: SDU.

Schuur, W.H. van (1984). *Structure in polical beliefs, a new unfolding model application to European party activists.* Amsterdam: CT Press.

Smets, P (1985). De school als organisatie. In J.L. Peschar & A.A. Wesselingh. *Onderwijssociologie.* Groningen: Wolters Noordhoff.

Sociaal-Cultureel Planbureau (1990). *School en schaal.* Rijswijk: CPB.

Swanborn, P.G. (1987). *Methoden van sociaal-wetenschappelijk onderzoek.* Meppel:Boom

Thoolen, E. (1987). Afvloeiingsregelingen in het onderwijs. *Meso, 35* (7), 7-18.

Toenders, L. (1989). Scholen in het informatietijdperk. In R. Van Hoewijk (ed.). *Schoolleider van vandaag, klaar voor morgen?: Liber amicorum voor Kees Gielen.* (pp. 15-47) Arnhem: Hogeschool Interstudie, Sector Management.

Versloot, A.M. (1990). *Ouders en vrijheid van onderwijs: Schoolkeuze in de provincie Utrecht.* Utrecht (dissertatie).

Visscher, A.J., & Amelsvoort, H.W.C.H. van (1993). *Evaluatie strategische factor personeelsbeleid BVE-sector (subsidieaanvraag SVO 93125).* Enschede: Universiteit Twente, OCTO.

Vries, A. de (1987). Kiezen voor een school. Schoolkeuze en marketing; een analyse van de 'afnemers' in het onderwijs. *Meso,* 22-26.

Wansink, H. (1992). *Een school om voor te kiezen, naar een actuele onderwijspolitiek.* Amsterdam: Bert Bakker.

Wieringen, A.M.L. van (1987). Deregulering en kritieken op onderwijsbeleid: vermogen van scholen tot voeren van bestuurlijk beleid. In L. Genemans (red.). *Autonomie van scholen en deregulering* (pp. 41-56). Nijmegen: ITS.

Manuscript ontvangen 28-4-1994
Definitieve versie ontvangen 2-9-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 2, pp. 133-153

De taakinhoud, de taakomvang en de taakbelasting van schoolleiders in het basisonderwijs

W.G.R. Stoel
Universiteit Twente, OCTO, Enschede

ABSTRACT

In most European countries there is no initial professional training for principals in primary education. This is the more remarkable since principals are expected to fulfill quite a lot of different tasks and expectations are high. Quite a number of publications state what principals *should* do in order to their job properly, whereas only a few publications are available in which is described what principals *actually* do when performing their function. This article summarizes the details of an extensive survey in which the actual behaviour of headteachers was the object of investigation. To this end respondents were asked either to answer a questionnaire about their activities or to keep a diary for a week. The results show that headteachers on the average have a work-load of over 50 hours a week, spread over 94 different activities. Nearly 50% of their working-hours headteachers function as teachers. Comparativily little time is dedicated to educational development.

INLEIDING

Aan de schoolleider in het basisonderwijs wordt zowel in onderwijsbeleidsnotities als in onderwijswetenschappelijke publikaties een belangrijke sturende en initiërende rol toegedacht. Frequent wordt het gedrag van de schoolleider gezien als *de* bepalende factor bij het tot stand komen van kwalitatief goed onderwijs op een school. De verwachtingen ten aanzien van de invloed van de schoolleider op het onderwijs-ontwikkelingsproces op de scholen zijn derhalve hooggespannen.

Onder invloed van diverse beleidsmaatregelen zijn de taken van schoolleiders in het basisonderwijs steeds diverser en omvangrijker van aard geworden. Zo heeft de invoering van de Wet op het Basisonderwijs in 1985 reeds gezorgd voor een diversificatie en verzwaring van schoolleiderstaken. De geformuleerde doelstellingen bij de invoering van deze wet wijzen uit dat men veel verwacht van de leraren en met name van de schoolleiders in het basisonderwijs. Een andere beleidsmaatregel die heeft geleid tot een verbreding van schoolleiderstaken betreft de centrale rol die onderwijsinstellingen op zich dienen te nemen bij de invoering van onderwijsinnovaties. Het initiatief ligt heel nadrukkelijk bij de school. De schoolleider heeft bij het implementeren van innovaties de functie van 'intern begeleider'. Vervolgens zijn er ontwikkelingen die ten doel hebben de schaal en de autonomie van scholen te vergroten. En om het geheel te completeren wordt tenslotte van schoolleiders van basisscholen verwacht dat zij een belangrijke bijdrage leveren aan de terugdringing van de omvang en de groei van het speciaal onderwijs.

Om antwoord te krijgen op de vraag hoe realistisch deze hoge verwachtingen ten aanzien van schoolleiders zijn, is nagegaan wat de inhoud is van het *huidige* takenpakket van de gemiddelde schoolleider.

VRAAGSTELLING ONDERZOEK

In 1983 omschreef Gielen (1983) het werk van een schoolleider al als: "druk, bedrijvig en in hoge mate onvoorspelbaar. Het bestaat uit zeer veel activiteiten en gebeurtenissen van korte tot

* Adres: OCTO, Universiteit Twente, Postbus 217, 7500 AE Enschede.

zeer korte duur. (.....) Het totaal aan activiteiten is bovendien erg gevarieerd." Een concreet inzicht in de feitelijke taakinhoud, taakomvang en taakbelasting van huidige schoolleiders in het basisonderwijs ontbreekt echter vooralsnog. Wel is er theoretisch het een en ander voorgesteld over de wijze waarop schoolleiderstaken kunnen worden gecategoriseerd. Op basis van een indeling van Mintzberg (1980) is door Gielen een indeling gegeven van de taakinhoud van schoolleiders in tien taakgebieden. Het betrof hier met name de algemene managementtaken van de schoolleider. Gielen plaatste in navolging van Mintzberg de taken van schoolleiders onder een drietal algemene managementrollen (interpersonal roles; informational roles; decisional roles). Ook Coulson (1986) plaatst de taken van schoolleiders in een aantal rollen. Hij maakt echter een onderscheid tussen de "Executive roles" en de "Professional roles". Hierbij wordt met betrekking tot de algemene administratieve managementtaken (executive role) aangesloten bij de hiervoor genoemde drie roltypen van Mintzberg, maar deze wordt uitgebreid een aantal roltypen die hij plaatst binnen de professionele onderwijskundige leiderschapsdimensie (professional role). Hieronder zijn taken te rangschikken zoals het maken van onderwijskundig beleid, het evalueren van leerlingen en staf en dergelijke. Met name deze dimensie van onderwijskundig leiderschap heeft de laatste jaren sterk in de belangstelling gestaan.(vgl. Levine en Lezotte, 1990; Hallinger en Murphy, 1987; Van de Grift, 1987 en Scheerens, 1989).

Uit het voorgaande blijkt dat de schoolleider met veel verschillende taken te maken heeft. In een aantal onderzoeken zijn observaties van schoolleidergedrag gedaan (Martin en Willower, 1981; Kmetz en Willower, 1982; Morris e.a., 1982). Uit deze observaties komt het beeld van de schoolleider als een drukke manager naar voren. Het blijkt dat schoolleiders lange werkweken hebben (gemiddeld 52 uur). Schoolleiders zijn vaak bezig met kortdurende, onvoorspelbare taken die zeer verschillend van aard zijn. Opvallend is dat de schoolleiders het overgrote deel van hun tijd nodig hebben voor bestuurlijke en beheersmatige zaken.

Martin en Willower rapporteren dat schoolleiders 17% van hun tijd besteden aan onderwijskundige aangelegenheden op schoolniveau zoals coördinatie van onderwijsinhoudelijke ontwikkelingen, stimulering en begeleiding van leerkrachten, planning en implementatie van roosters. Verondersteld kan worden dat de schoolleider juist bij de uitvoering van deze taken, die tezamen het onderwijsbeleid vormen, invloed kan uitoefenen op prestaties en gedrag van leerlingen. Het feit dat schoolleiders hieraan zo weinig tijd besteden kan voor een deel verklaard worden uit de hoeveelheid tijd en aandacht die administratieve en bestuurlijke taken opeisen.

Een andere verklaring wordt gevonden in de structuur van schoolorganisaties. Het onderwijssysteem binnen schoolorganisaties wordt gekenmerkt door de autonomie van professionele leerkrachten (Emans, 1982). Leerkrachten zijn elders opgeleide 'professionals', waarvan verondersteld wordt dat ze hun onderwijsgevende taak zonder begeleiding zelfstandig kunnen uitvoeren. Omdat leerkrachten afzonderlijk bezig zijn in klassen, en klassesituaties zeer verschillend zijn, is het lesgeven een moeilijk stuurbaar en controleerbaar gebeuren. De autonomie van leerkrachten maakt dat schoolleiders weinig invloed kunnen en durven uitoefenen op de taakuitvoering van leerkrachten. Door middel van vaksecties en lerarenvergadering wordt de noodzakelijke coördinatie verkregen, maar een goed ondersteunend onderwijsbeleid dat richting en sturing aan leerkrachten geeft, is op veel scholen niet aanwezig. In de praktijk blijkt dat sturende en begeleidende activiteiten, die in andere, vergelijkbare, organisaties wel voorkomen, weinig door de schoolleiding plaatsvinden. Activiteiten als het voeren van functioneringsgesprekken, het doen van klasbezoeken en het bespreken van onderwijskundige zaken in vergaderingen worden weinig gedaan. Zoals eerder is aangegeven zou de schoolleider juist door middel van zijn onderwijstechnische taak invloed kunnen hebben op het functioneren van leerkrachten en leerlingen. Uit de effectieve scholenstudies blijkt dat uitvoering van onderwijstechnische zaken door de schoolleider van invloed is op de schooleffectiviteit.

Brookover e.a. (1979) rapporteren resultaten verkregen door middel van case-studies waarin twee effectieve en twee niet-effectieve scholen vergeleken werden. Uit de resultaten concluderen zij dat schoolleiderschap op effectieve scholen duidelijk verschilt van die van niet-effectieve scholen. Op de effectieve scholen werd de schoolleider gezien als een uitstekend onderwijs-

kundig leider. Deze schoolleiders waren op organisatorisch en bestuurlijk gebied goede leiders en vervulden op onderwijskundig gebied een actieve, stimulerende rol. Zij hadden regelmatig bijeenkomsten met leraren, waarin de prestaties van de leerlingen besproken werden en probeerden tijdens deze bijeenkomsten de leraren ervan te overtuigen dat de prestaties van de leerlingen verbeterd konden worden. Men bezocht regelmatig de lessen om de leraren suggesties te geven voor het verbeteren van de onderwijsvaardigheden.

Ook stimuleerden deze schoolleiders de leraren tot het bezoeken van seminars, workshops en inservice-trainingen. De schoolleiders van niet-effectieve scholen waren goede organisatorische en bestuurlijke leiders. Zij besteedden hun tijd hoofdzakelijk aan administratieve, organisatorische zaken en discipline problemen. Zij toonden echter weinig aandacht voor de prestaties van leerlingen. Ook konden zij de leraren niet stimuleren tot verbetering van onderwijsvaardigheden en verbetering van leerlingprestaties.

Samenvattend wordt geconcludeerd dat het belangrijkste verschil tussen schoolleiders is dat schoolleiders op effectieve scholen naast hun administratieve en organisatorische taken op een aantal manieren veel aandacht besteden aan onderwijskundige taken.

In een onderzoek van Wellisch, Mac Queen, Carriere en Duck (1978) werd het schoolleiderschap op 9 effectieve en 13 niet-effectieve lagere scholen vergeleken. Gedurende een periode van twee jaar werden effectieve scholen van niet-effectieve scholen onderscheiden op basis van vooruitgang van de leerlingen in de vakken lezen en rekenen. Zowel bij schoolleiders als leraren werd een enquête over schoolleiderschap afgenomen. Uit de resultaten van beide enquêtes blijkt dat het schoolleiderschap van effectieve en niet-effectieve scholen op een aantal punten verschilt. Ten eerste blijkt dat schoolleiders van effectieve scholen meer belangstelling hebben voor onderwijskundige zaken, een onderwijskundige visie hebben en deze visie overbrengen op de leraren.

Ten tweede blijkt dat schoolleiders van effectieve scholen de communicatie over onderwijskundige zaken bevorderen door het regelmatig houden van discussies over en evaluaties van het werk van de leraren. Ten derde participeren deze schoolleiders in beslissingen over selectie van lesmateriaal, planning en evaluatie van schoolprogramma's. Ten vierde zorgen deze schoolleiders voor een goede coördinatie van lesprogramma's.

Andere onderzoeken leveren vergelijkbare conclusies op als die welke hiervoor besproken zijn (vgl. bijv. Rutter e.a., 1979 en Edmonds, 1979). De gemeenschappelijke conclusie is dat schoolleiders naast hun organisatorische en bestuurlijke taken veel aandacht besteden aan het bevorderen van de onderwijspraktijk. Dit aspect van schoolmanagement wordt onderwijskundig leiderschap (instructional leadership) genoemd.

Er is met betrekking tot de feitelijke taakinvulling door de schoolleider wat betreft algemene en onderwijskundige managementtaken vooralsnog weinig bekend. Het voorgestelde onderzoek zal deze lacune opheffen en er zal zowel aandacht worden besteed aan de feitelijke, als aan de wenselijk geachte taken van schoolleiders in het kader van de voorgestelde beleidsmaatregelen die leiden tot een vergroting van de autonomie van scholen en tot schaalvergroting.

In de inleiding is reeds beknopt ingegaan op de aanleiding tot, de achtergrond van en de vraagstelling van het onderzoek. De vier hoofdvraagstellingen van onderzoek kunnen als volgt worden geformuleerd:

1) Welke schoolinterne zaken worden door schoolleiders uitgevoerd en met welk doel? Hoeveel tijd wordt door schoolleiders aan deze taken besteed? Welke taken worden gedelegeerd en hoe verloopt de controle en de uitvoering van die taken?
2) Welke schoolexterne zaken worden door schoolleiders uitgevoerd en met welk doel? Wat is de aard en de frequentie van externe contacten, hoeveel tijd is gemoeid met het onderhouden van deze contacten en wat zijn de consequenties van deze contacten voor de eigen beslissingsruimte en hoe beoordelen schoolleiders het belang van het onderhouden van deze externe contacten?
3) Welke consequenties zal de door de overheid voorgestane schaal- en autonomievergroting

hebben voor het schoolmanagement? In hoeverre komen de taken betreffende het onderwijskundig management in de school gevaar en welke oorzaken zijn daarvoor aan te wijzen?
4) Welke taken (moeten) naar het oordeel van direct betrokkenen en deskundigen, worden gerekend tot die van het bevoegd gezag en welke tot die van de schoolleider en waarom?

Alleen de eerste twee onderzoeksvragen zullen in dit artikel worden beantwoord. Deze eerste twee vragen behoeven een empirisch onderzoek naar het soort taken dat de schoolleider uitvoert; de tijdsbesteding aan deze taken; de gepercipieerde taakbelasting, de mate van delegatie van taken of taaksoorten aan anderen binnen de schoolorganisatie en de verdeling van taken tussen het bevoegd gezag en het schoolmanagement. Het betreft hier zowel taken die schoolintern (bijv. personeelsmanagement) als taken die schoolextern zijn (oudercontacten, netwerken e.d.) De taken hebben betrekking op verschillende functies die de schoolleider kan vervullen. De gemiddelde schoolleider heeft naast zijn verantwoordelijkheden als groepsleerkracht, ook de verantwoordelijkheid voor het onderwijskundig beleid, het personeelsbeleid, het beheer en de financiën, het onderhouden van externe contacten, zowel op professioneel gebied (Voortgezet Onderwijs; Onderwijsvoorrangsbeleid; Onderwijsbegeleidingsdiensten e.d.) als met de ouders, alsmede het verzorgen van public relations en activiteiten die verband houden met schoolmarketing (Gielen en Kraaykamp, 1981). In hun analyse van de schoolleidersfunctie spreken deze auteurs in dit verband tevens van taken en activiteiten in het interne veld (team en bestuur), pedagogisch veld (ouders en leerlingen) en het maatschappelijk veld (schoolomgeving). Het is voor veel schoolleiders ondoenlijk om in de betrekkelijk geringe tijd die hun per dag of week ter beschikking staat voor het uitvoeren van (onderwijskundige) managementtaken om alle taken die hun worden toebedeeld uit te voeren. Veel schoolleiders ervaren het omgaan met hun tijd als een groot probleem (Dolmans, 1990). Er kan sprake zijn van een *kwantitatief-* (een te grote taakomvang voor het aantal werkuren), een *kwalitatief-* (een onevenwichtige verdeling van de tijdsinvestering over de taken) en een *subjectief-* (een alszodanig ervaren) tijdsprobleem. Het delegeren van bepaalde taken is dan ook noodzakelijk. Het is daarnaast nog de vraag of alle voornoemde taken wel tot het terrein van de schoolleider zouden moeten behoren, of dat er een goede en efficiënte taakverdeling tussen bevoegd gezag, schoolleiding en overige teamleden nodig is.

ONDERZOEKSOPZET

Er is een instrumentarium ontwikkeld om de concrete activiteiten van schoolleiders in kaart te brengen. Nadat het instrumentarium was ontwikkeld, is dit uitgetest op zijn bruikbaarheid en geldigheid in een pilotstudie en in een try-out-studie. Vervolgens is dit instrumentarium gebruikt in een onderzoek waaraan uiteindelijk 177 schoolleiders van basisscholen hun medewerking hebben verleend.

Ontwikkeling instrumenten

In de eerste twee onderzoeksvragen staan drie begrippen centraal. Deze drie begrippen zijn, taakinhoud, taakomvang en taakbelasting. De taakinhoud heeft betrekking op de beschrijving van de afzonderlijke taken welke door de schoolleider worden uitgevoerd. Een belangrijk deel van het onderzoek bestond dan ook uit een instrumentconstructie. Daarbij is getracht de drie centrale begrippen in dit onderzoek te operationaliseren. In de eerste plaats is een inventarisatie gemaakt van alle taken die door schoolleiders zouden kunnen worden uitgevoerd; een inventarisatie van de taak*inhoud* derhalve. Bij deze inventarisatie is gebruik gemaakt van relevante literatuur en van de inbreng van de geïnterviewde schoolleiders. In totaal zijn 94 verschillende activiteiten van schoolleiders onderscheiden. Het tweede centrale begrip betreft de taak*omvang*. Deze taakomvang is in kaart gebracht door schoolleiders per onderscheiden activiteit te vragen hoeveel minuten zij per, zelf te kiezen frequentie: dag, week, maand, etc., aan deze activiteit besteden. Als derde begrip is de taak*belasting* geoperationaliseerd. Het betreft hier een catego-

rie die rechtstreeks in verband kan worden gebracht met de subjectief beleefde taakbelasting van schoolleiders. Deze ervaren taakbelasting is geoperationaliseerd door per onderscheiden activiteit te vragen in hoeverre de schoolleider naar zijn eigen mening in voldoende mate tijd heeft geïnvesteerd om de desbetreffende activiteit naar behoren te kunnen uitvoeren.

Er is gekozen voor een werkwijze waarbij via twee verschillende methoden is geïnventariseerd wat de taakinhoud, de taakomvang en de taakbelasting is van schoolleiders. Het meest voor de hand liggende instrument om de activiteiten van de schoolleider en de tijd die hij eraan besteedt in kaart te brengen is uiteraard het tijdschrijfformulier. Dit is dan ook in het onderzoek gebruikt. Er is een logboek ontwikkeld waarin de schoolleider gevraagd is om een week lang van minuut tot minuut bij te houden wat hij/zij doet. Naast het tijdschrijfsysteem is in het onderzoek een tweede manier van registreren gebruikt. Met behulp van een vragenlijst is de schoolleider gevraagd hoeveel tijd hij gemiddeld, naar eigen zeggen besteedt aan de verschillende onderscheiden activiteiten. De te hanteren tijdseenheid dag, week, maand, kwartaal of jaar kon de respondent zelf kiezen.

In zowel de vragenlijst als het logboek is verder expliciet gevraagd welke activiteiten op schoolniveau door de schoolleider worden gedelegeerd aan andere leden van het schoolteam dan de adjunct-directeur. Daarbij is tevens nagegaan hoeveel tijd deze andere leden van het team naar de mening van de schoolleider besteden aan deze gedelegeerde activiteiten.

Procedures

De vragenlijsten en logboeken zijn verzonden aan in totaal 600 a-select gekozen basisscholen. De helft van deze 600 scholen heeft – verdeeld over drie verschillende tijdstippen – de vragenlijsten ontvangen; de andere helft de logboeken, eveneens verdeeld over drie tijdstippen. Ook ten behoeve van de adjuncten is een vragenlijst dan wel logboek verzonden. Overigens waren niet alle aangeschreven scholen groot genoeg om een adjunct te hebben. De 300 logboeken zijn in drie oplagen van elk honderd op drie momenten in het schooljaar 1992/1993 verzonden: in oktober en november 1992 en in januari 1993. In totaal 80 scholen retourneerden een ingevuld en bruikbaar logboek; dit komt overeen met 28% van de benaderde scholen. Van de schoolleiders die een vragenlijst ontvingen retourneerden 97 een ingevuld exemplaar, hetgeen overeenkomt met een respons van 34%.

De versie ten behoeve van de schoolleider en zijn adjunct verschilden op een aantal aspecten. Zo is alleen aan de schoolleider gevraagd om informatie te verstrekken over een aantal achtergrondkenmerken van de school en het schoolteam en welke taken op school door de schoolleider worden gedelegeerd aan anderen dan aan de adjunct. Deze vragen zijn niet (nogmaals) gesteld aan de adjunct; voor het overige waren de beide versies inhoudelijk gelijk.

Kenmerken van de responderende scholen

Van de responderende scholen is informatie gevraagd over een aantal achtergrondkenmerken van de school en de leerlingen- en docentenpopulatie. De verzameling van deze informatie diende een tweeledig doel. In de eerste plaats werd het daardoor mogelijk om de kenmerken van de responderende scholen te vergelijken met de kenmerken van scholen zoals deze voor de populatie scholen als geheel bekend zijn. In de tweede plaats konden op deze wijze de achtergrondkenmerken gerelateerd worden aan de gegeven antwoorden met betrekking tot de taakinhoud, de taakomvang en de taakbelasting van de schoolleiders. In de volgende alinea's worden de kenmerken van de responderende scholen beschreven en vergeleken met de beschikbare informatie op dit punt voor de populatie als geheel. In een van de volgende paragrafen wordt gerapporteerd over de al dan niet gevonden relaties tussen deze achtergrondkenmerken en de taakomvang en de taakbelasting van de schoolleiders.

Van de scholen was 28% van protestants-christelijke signatuur, ruim 33% van rooms-katholieke signatuur, 28% van de scholen verzorgden openbaar onderwijs en 5% was van algemeen bijzondere signatuur.

Qua plaats van vestiging van de responderende scholen blijkt dat ruim 78% afkomstig is uit plaatsen met minder 50.000 inwoners, 7,5% van de scholen is gevestigd in plaatsen met meer

dan 50.000 maar minder dan 100.000 inwoners, 7,5% is gevestigd in plaatsen met minder dan 400.000 maar meer dan 100.000 inwoners en bijna 7% van de responderende scholen komt uit plaatsen met meer dan 400.000 inwoners. Van de responderende schoolleiders is bijna 17% vrouw en ruim 83% man. Twee respondenten gaven geen antwoord op deze vraag. Landelijk ligt het percentage vrouwelijke schoolleiders op 12,8% (CBS, zakboek onderwijsstatistieken, 1992; cijfers 1990, pag. 93). Het gemiddeld aantal leerlingen op de responderende (vragenlijst)scholen was 171 met een minimum van 22 en een maximum van 412. Volgens het zakboek onderwijsstatistieken van het CBS uit 1992 (pag. 24), was het gemiddeld aantal leerlingen per school in 1990 eveneens 171.

De *gemiddelde* leeftijd van de schoolteams was 38,7 jaar met een minimum van 26 en een maximum van 49 jaar. Hoe dit in de populatie ligt is onbekend. Het percentage zgn."1.9"-leerlingen was 6,5% met een maximum van bijna 72% en een minimum van nul procent. De responderende schoolleiders hadden gemiddeld bijna 14 jaar ervaring als schoolleider, met een minimum van 1 en een maximum van 32 jaar. Daarnaast hadden zij gemiddeld genomen 21 jaar ervaring in het basisonderwijs met een minimum van 11 jaar en een maximum van 37 jaar. Ruim 32 procent van de scholen had de beschikking over een conciërge. Hoe dit in de populatie ligt is onbekend.

De verschillen tussen "vragenlijst-" en "logboek"-scholen zijn qua achtergrondkenmerken marginaal te noemen. De gemiddelde schoolgrootte van de "logboek"-scholen ligt wat hoger, evenals, logischerwijze, het gemiddeld aantal personeelsleden. Het percentage vrouwelijke schoolleiders lag bij degenen die het logboek bijhielden wat lager (14%) dan bij degenen die een vragenlijst invulden (17%). Op de "logboek"-scholen blijkt evenwel vaker een conciërge aanwezig te zijn dan op de "vragenlijst"-scholen; ook het percentage "1.9"-leerlingen verschilt voor beide groepen.

De verschillen tussen beide groepen respondenten zijn beperkt; de invloed die deze verschillen wellicht hebben op mogelijke verschillen in scores tussen "logboek"-scholen en "vragenlijst"-scholen kan niet empirisch vastgesteld worden vanwege het samengaan van deze met het verschil in methode van gegevensverzameling. Wat men echter wel kan verwachten als gevolg van het feit dat "logboek"-scholen gemiddeld groter zijn dan "vragenlijst"-scholen is dat schoolleiders van "logboek"-scholen, over wat meer ambulante lestijden zullen beschikken dan hun collega's van de "vragenlijst"-scholen. Het gevolg daarvan is weer dat deze laatsten wat meer tijd zullen besteden aan onderwijs verzorgen.

Wat betreft de gemiddelde waarden op de achtergrondkenmerken van de responderende scholen en schoolleiders is er geen reden om te veronderstellen dat deze afwijken van de waarden die in de populatie gelden. Dit geldt uiteraard alleen voor zover van deze laatste populatiewaarden bekend zijn.

Gegevensbewerking en gegevensverwerking

De ruwe gegevens, zoals verkregen door middel van de vragenlijsten, zijn via de geëigende methoden gecodeerd en ingevoerd. De gegevens verkregen via de logboeken vereisten ten dele vooraf een bewerking. Alle activiteiten die de schoolleiders noemden werden door twee personen onafhankelijk van elkaar gecodeerd op basis van de codes die waren toegekend aan de vragen in de vragenlijst; een activiteit in het logboek werd dus dezelfde code toegekend als deze activiteit in de vragenlijst had. Wanneer consensus ontbrak werd door een derde beoordelaar eerst een codering toegekend en daarna werd in onderling overleg besloten voor welke codering men koos.

Wanneer een activiteit uit de logboeken in de vragenlijst niet voorkwam, werd deze ondergebracht in een "rest"-categorie binnen één van de vijf onderscheiden hoofdcategorieën in de vragenlijst. Daarna werden de logboekgegevens op dezelfde manier verwerkt als de gegevens afkomstig van de vragenlijsten. De gedelegeerde taken die zijn genoemd in zowel logboek als vragenlijst zijn op eenzelfde manier bewerkt.

De gegevensverwerking bestond in eerste instantie uit het maken van rechte tellingen (vgl. Stoel, 1994). Daarnaast heeft een clustering van de 94 onderscheiden activiteiten op inhoudelijke gronden plaatsgevonden teneinde een meer toegankelijke beschrijving van de resultaten mogelijk te maken. De vier clusters die in eerste instantie werden onderscheiden zijn: A. Alle activiteiten die gerelateerd zijn aan het *onderwijs verzorgen* door de schoolleider zelf, zoals lesgeven, correctie-werkzaamheden, voorbereiding van lessen e.d.; B. Alle activiteiten die onder te brengen waren in het cluster *vergaderen en overleggen*; C. De activiteiten die de schoolleider in principe achter zijn bureau kan verrichten zoals leerlingadministratie, financiële zaken afwikkelen e.d. en tenslotte een cluster D. waarin activiteiten van diverse aard zijn ondergebracht, zoals bij- en nascholingsactiviteiten, conciërge-activiteiten en activiteiten die samenhangen met toezicht en beheer.

Ook zijn correlationele analyses uitgevoerd. Nagegaan is of er relaties konden worden aangetoond tussen de tijd die besteed werd aan afzonderlijke (en geclusterde) activiteiten en achtergrondkenmerken van de school, de schoolleider en het schoolteam. Deze samenhangen zijn nagegaan voor de tijdsbesteding per activiteit van zowel de schoolleider als zijn adjunct. Deze correlaties zijn ook berekend voor de gedelegeerde taken in zowel logboek als vragenlijst. Alle berekeningen zijn voor logboek- en vragenlijstgegevens apart uitgevoerd. Een aantal contextvariabelen waren discreet van aard; in die gevallen zijn variantie-analyses uitgevoerd in plaats van correlaties berekend.

RESULTATEN

In deze paragraaf worden de resultaten verkregen via logboek en vragenlijst gelijktijdig beschreven en wordt ingegaan op geconstateerde overeenkomsten en verschillen, zodat een beeld verkregen wordt van de (on)vergelijkbaarheid van de via deze twee methoden verkregen informatie.

Taakinhoud en taakomvang van de schoolleider

Een aantal relativerende opmerkingen vooraf is op zijn plaats. In de eerste plaats dient men te bedenken dat de respondenten zelf rapporteren hoeveel tijd zij, gemiddeld, per tijdseenheid gebruiken voor een groot aantal verschillende activiteiten. Gezien de vraagstelling en het belang van het onderzoek voor betrokken respondenten is het niet ondenkbaar dat er, al dan niet bewust, (hier en daar) door (sommige) respondenten "naar boven afgerond" wordt bij de beantwoording van de vragen (één schoolleider heeft bijvoorbeeld aangegeven dat hij of zij gemiddeld(!) bijna 90 uur per week werkt. Daar staat tegenover dat veel zeer druk bezette potentiële respondenten zeggen vanwege deze drukte niet in het onderzoek te kunnen participeren. Dat zou tot gevolg kunnen hebben gehad dat er naar verhouding minder druk bezette schoolleiders hebben geparticipeerd in het onderzoek. Leidt het werkzaam zijn van het eerste mechanisme wellicht tot een overschatting van de in feite bestede tijd; het werkzaam zijn van het tweede mogelijk werkzame mechanisme zou tot een onderschatting aanleiding kunnen geven.

Teneinde onder andere op dit punt een kruisvalidatie van resultaten mogelijk te maken is ook via een, voor de respondenten zeer arbeidsintensieve, logboek-methode informatie verzameld over de tijdsbesteding van schoolleiders. Aangezien de verschillen tussen de resultaten verkregen via de vragenlijsten en die verkregen via de logboeken minimaal blijken te zijn, is in de tabellen 1 t/m 5 volstaan met het weergeven van de resultaten verkregen via de vragenlijsten. De geringe verschillen die wel blijken te bestaan tussen de resultaten verkregen via beide methoden van gegevensverzameling worden beschreven aan het einde van deze paragraaf.

Tenslotte dient nog te worden vermeld dat de vragenlijst-methode van gegevensverzameling aanleiding kan geven tot het "dubbel declareren" van werkzaamheden. Wanneer een schoolleider bijvoorbeeld "onder de les" andere in de vragenlijst opgenomen activiteiten verricht, zoals bijvoorbeeld correctiewerk, ad hoc-gesprekken met ouders e.d. leidt dit automatisch ook tot een overschatting van de werkelijk bestede tijd.

Tabel 1. Gemiddelde, door de schoolleider, bestede tijd (in minuten per week) en standaarddeviatie per onderscheiden activiteit op hoofdaspecten.

Totaal bestede tijd aan activiteit	Gemiddeld aantal minuten	Standaard deviatie	Minimum	Maximum
Overleggen	586	211	191	1124
Lesgeven (+ voor- en nawerk)	1325	528	60	2707
Bureau-activiteiten	670	343	179	2018
Toezicht en beheer	295	166	24	815
Diverse activiteiten	138	34	0	245

In deze paragraaf worden de door respondenten aangegeven tijden integraal weergegeven en, waar aan de orde, gesommeerd. Eerst wordt echter in tabel 1 een overzicht gegeven van de gemiddelde taakomvang van de responderende schoolleiders op hoofdaspecten. De vijf aspecten die zijn onderscheiden zijn: overleggen, onderwijs verzorgen, bureau-activiteiten, toezicht en beheer en een rubriek diverse activiteiten waaronder vallen nascholings- en extra-curriculaire activiteiten.

De spreiding in deze gemiddelden is bijzonder groot; de meest tijdseffectief werkende fulltime aangestelde schoolleider zegt een weektaak te hebben van 26 uur, terwijl de schoolleider die de meeste tijd zegt te besteden aan een gemiddelde weektaak komt van bijna 90 uur. Deze grote verschillen qua tijdsbesteding tussen schoolleiders en ook tussen de diverse adjuncten is minstens zo betekenisvol als de gegevens die beschikbaar komen via de inventarisatie van de gemiddeld bestede tijd per activiteit.

De hierna weergegeven conclusies betreffen de *gemiddelde* tijdsbesteding per week van de *gemiddelde* schoolleider. Ter wille van de leesbaarheid is ervan afgezien om telkenmale te melden dat de gepresenteerde scores gemiddelden betreffen.

De schoolleider blijkt iets meer dan 50 uur per week te besteden aan zijn taken als schoolleider. Zoals te verwachten was wordt, naar verhouding, de meeste tijd daarvan besteed aan lesgeven en aan daarmee samenhangengende activiteiten als lesvoorbereiding en correctiewerkzaamheden; iets minder dan de helft van de tijd van de schoolleider. Bureau-activiteiten, zoals financiële zaken, postafhandeling e.d. vragen gemiddeld bijna 11 uur per week, terwijl bijna 10 uur per week in overlegsituaties wordt doorgebracht. In de volgende alinea's wordt aangegeven aan welke activiteiten binnen elk van deze hoofdcategoriecën de tijd wordt besteed.

In *tabel* 2 wordt gedetailleerd weergegeven in welke overlegsituaties de schoolleider participeert en hoeveel tijd hij per week aan elk hiervan besteedt. Het betreft hier een beeld van de *gemiddelde* schoolleider (die dus niet daadwerkelijk bestaat); het zal dus niet zo zijn dat iedere schoolleider alle hier genoemde activiteiten uitvoert.

Daar waar als minimum tijdsbesteding een nul is vermeld is dit een indicatie van het feit dat tenminste één schoolleider geen tijd zegt te besteden aan deze activiteit. Het overzicht zoals weergegeven in tabel 2 spreekt voor zichzelf; volstaan wordt derhalve met het noemen van een aantal trends of opvallende zaken.

Door de responderende schoolleiders wordt bijna 10 uur per week besteed aan overleggen, besprekingen en vergaderingen. De spreiding daarin is echter groot; sommige schoolleiders besteden aan deze activiteiten iets meer dan drie uur per week terwijl andere schoolleiders bijna 19 uur per week overleggen, vergaderen en bespreken. Het blijkt dat er binnen de categorie "overleggen, bespreken en vergaderen" drie overlegsituaties zijn die relatief de meeste tijd van de schoolleider vragen: overleg met het team als geheel, overleg met individuele docenten of met groepjes docenten en overleg met ouders. Het betreft hier echter *gemiddelden*, de verschillen tussen scholen en schoolleiders zijn bijzonder groot.

Omdat de tijd die wordt besteed aan overleggen op zich niets zegt over de inhoud ervan, is voor een van de overlegsituaties, het teamoverleg, nagegaan welke onderwerpen daarin aan de

Tabel 2. Gemiddelde, door de schoolleider, bestede tijd (in minuten per week) en standaarddeviatie per onderscheiden activiteit binnen de categorie "overleggen, bespreken en vergaderen".

Totaal bestede tijd aan activiteit	Gemiddeld aantal minuten	Standaard deviatie	Minimum	Maximum
Overleg met het docententeam	103	40	10	300
Overleg met adjunct- of plv. directeur	62	79	0	600
Overleg met docenten: afgesproken	29	37	0	270
Overleg met docenten: onafgesproken	68	60	0	338
Functioneringsgesprek met docent	10	16	0	90
10-minuten gesprekken met ouders	20	16	0	72
Algemene ouderavond	8	7	0	60
Overleg met ouders op afspraak	30	34	2	240
Overleg met ouders niet op afspraak	41	37	0	153
Overleg met ouderraad	22	24	0	167
Overleg met medezeggenschapsraad	14	11	0	56
Overleg met overige groepen ouders	11	15	0	90
Overleg met het bevoegd gezag	36	22	2	120
Overleg met schoolbegeleiders	24	29	2	180
Overleg met politie etc.	6	9	0	65
Overleg met collega's basisonderwijs	20	17	0	120
Overleg collega's speciaal onderwijs	6	9	0	44
Overleg collega's voortgezet onderwijs	5	4	0	28
Overleg met de buurt	3	8	0	60
Overleg met O&W	5	1	0	7
Overleg met onderwijsinspectie	6	27	0	270
Overleg leveranciers schoolmateriaal	12	13	0	75
Overleg met vakleerkrachten	7	8	0	30
Overleg met de muziekconsulent	1	4	0	28
Overleg met de Pabo	4	9	0	60
Overleg met de school(tand)arts	3	4	0	23
Overleg met onderzoekers	3	4	0	28
Overleg met vakorganisaties	2	6	0	48
Overleg met pedagogische centra	1	2	0	14
Overleg met de gemeente	8	11	0	60
Overleg met de kerk	5	7	0	48
Participatie in landelijke projecten	22	36	0	203

orde komen en welk deel van de tijd die aan het teamoverleg wordt besteed, wordt gewijd aan welke onderwerpen. De meeste tijd van het teamoverleg (ongeveer eenderde deel van het totaal) wordt besteed aan onderwijsinhoudelijke zaken. Een goede "tweede" is het bespreken van aangelegenheden van individuele leerlingen en/of hun ouders (13% van de tijd die wordt besteed aan teamoverleg). Ongeveer 10% van de tijd wordt zowel besteed aan de regel- en rooster-aangelegenheden als het bespreken van het onderwijsbeleid in zijn algemeenheid, als aan terugrapportage van inhouden van vergaderingen waar niet het (gehele) team bij aanwezig was.

De responderende schoolleiders geven aan gemiddeld 5% van het teamoverleg te besteden aan "andere" dan de voorgestructureerde activiteiten. Voor zover deze andere activiteiten werden gespecificeerd kon daaruit worden afgeleid dat dit met name de volgende zaken betrof: (organisatie van) sportdagen, schoolreizen, festiviteiten, vieringen en andere buitenschoolse activiteiten.

Uit de gegevens zoals weergegeven in Tabel 2 blijkt dat er gemiddeld per week een uur en drie kwartier wordt uitgetrokken voor teamoverleg. Bijna eenderde deel daarvan wordt, gezien hetgeen hiervoor werd weergegeven, besteed aan onderwijsinhoudelijke zaken die de school als geheel betreffen. En ruim 40% wanneer men het bespreken van het onderwijsbeleid in zijn

Tabel 3. Gemiddelde, door de schoolleider, bestede tijd (in minuten per week) en standaarddeviatie per onderscheiden activiteit binnen de categorie "les geven".

Activiteit	Gemiddelde	Standaard deviatie	Minimum	Maximum
Lesgeven	867	323	0	1395
Afstemmingsoverleg met vervanger	28	25	0	135
Voorbereiding onderwijs eigen groep	155	139	0	675
Correctie werk leerlingen	197	156	0	540
Verzorgen remediële activiteiten	43	67	0	300
Onderwijs geven aan ouders	1	12	0	120
Begeleiding van Pabo-stagiaires	12	21	0	120
Begeleiden ouders met een taak	6	9	0	45
Voorbereiden 10-minuten-gesprekken	12	16	0	90
Voorbereiding vieringen eigen groep	11	11	0	60

algemeenheid daarbij betrekt. In een gemiddeld, wekelijks teamoverleg op een gemiddelde school wordt derhalve iets meer dan 30 minuten uitgetrokken voor het bespreken van (de ontwikkeling van) het onderwijs-inhoudelijke beleid op school. De inhoud van een gemiddeld teamoverleg lijkt een goed afspiegeling te zijn van het, zeer heterogene, takenpakket van de schoolleider: veel onderwerpen van zeer diverse aard passeren de revue.

In het kader van de beleidsinspanningen om het basisonderwijs en het speciaal onderwijs wat meer op elkaar te betrekken met als uiteindelijk doel de groei en ook de omvang van het speciaal onderwijs te verminderen is het opvallend dat er bijzonder weinig tijd wordt geïnvesteerd in overleg met collega's uit het speciaal onderwijs. Wanneer men de lijst zoals weergegeven in tabel 2 overziet, moet men tot de conclusie komen dat er betrekkelijk weinig tijd wordt geïnvesteerd in de werkelijk externe contacten van scholen. De schoolleider verzorgt in het merendeel van de gevallen ook onderwijs. Ook in 1993 blijkt de schoolleider in de eerste plaats docent te zijn en besteedt hij aan het geven van onderwijs aan een eigen groep het merendeel van zijn totale werktijd.

Opvallend in dit geheel is de tijd die de schoolleider zegt te besteden aan lesvoorbereiding en correctie-activiteiten: bijna 6 uur per week. Hoeveel van deze activiteiten binnen de reguliere lestijd (werk nakijken als de klas zelfstandig werkt) plaatsvinden is niet te zeggen op grond van deze onderzoeksgegevens. Daarnaast wordt door de schoolleider, relatief gesproken, vrij veel tijd uitgetrokken voor remediële activiteiten in verband met leerlingen uit andere groepen dan de zijne, namelijk ongeveer drie kwartier per week. Een enkeling zegt zelfs vijf uur per week voor deze activiteiten uit te trekken.

Ten aanzien van het verzorgen van onderwijs worden eveneens grote verschillen tussen schoolleiders geconstateerd die uiteraard voor een groot deel zijn te herleiden tot verschillen in mate van ambulant zijn van de schoolleider: sommige schoolleiders zeggen echter zowel aan correctiewerk als aan lesvoorbereidende werkzaamheden meer dan negen uur per week te besteden.

Ook met betrekking tot de vervulling van de zgn. "bureau-activiteiten" is de diversiteit aan activiteiten groot; van de schoolleider wordt verwacht dat hij zeer veel verschillende taken vervult. In de vragenlijst zelf zijn 28 bureau-activiteiten onderscheiden.

Het doornemen van de post vraagt kennelijk veel tijd. Gemiddeld wordt voor deze activiteit twee uur per week uitgetrokken, met een uitschieter van 9 uur per week. Ook zijn er schoolleiders die kennelijk aan 10 minuten per week genoeg hebben.

De teamvergaderingen zelf leggen, qua tijdsinvestering, zo bleek reeds eerder, een relatief groot beslag op de schoolleider, maar ook de voorbereiding van deze vergaderingen vraagt nog eens een uur per week. Uit deze tabel blijkt dat de schoolleider per week ongeveer anderhalve dag per week bureau-activiteiten verricht. De op dit punt minst belaste schoolleider besteedt drie uur per

Tabel 4. Gemiddelde, door de schoolleider, bestede tijd (in minuten per week) en standaarddeviatie per onderscheiden activiteit binnen de categorie "bureau-activiteiten".

Totaal bestede tijd aan activiteit	Gemiddeld aantal minuten	Standaard deviatie	Minimum	Maximum
Inkomende post doornemen	119	78	10	540
Correspondentie met O&W	1	3	0	14
Correspondentie met inspectie	15	17	0	100
Correspondentie met ouders	13	18	0	90
Corr. met maatschappelijke instellingen	9	13	0	60
Corr. met leveranciers schoolmaterialen	8	9	0	60
Corr. met leveranciers diensten	14	23	0	120
Overige correspondentie	48	55	0	270
Voorbereiden van teamvergaderingen	63	44	8	270
Voorbereiden vergaderingen ouderraad	8	9	0	60
Voorbereiden overleg inspectie	2	3	0	28
Voorbereiden overleg bevoegd gezag	13	12	0	69
Voorbereiden overleg onderst. instanties	10	14	0	111
Afhandeling financiële zaken	50	47	0	270
Voorbereiden incidenteel overleg ouders	9	12	0	60
Bijhouden van leerlingdossiers	20	26	0	139
Vormgeving onderwijsleerproces	63	81	0	600
Activiteiten ivm leerlingadministratie	30	40	0	270
Activiteiten ivm personeelsadministratie	26	28	0	180
Bestuderen brochures e.d. van O&W	44	36	0	240
Verdeling lesoverstijgende taken	8	12	0	60
Regelen vervangende leerkrachten	13	19	0	120
Indelen groepen en leerkrachten	7	14	0	120
Bijhouden van schoolarchief	16	21	0	120
Meewerken aan onderzoek	7	5	1	23
Opstellen van activiteitenplan	15	16	0	60
Schrijven van info-bulletins	33	46	0	293
Organisatie vieringen andere groepen	7	9	0	43

week aan dergelijke activiteiten; de meest belaste schoolleider ruim 36 uur per week. Opvallend is dat er naar verhouding weinig tijd wordt besteed aan (of overblijft voor?) de vormgeving van het onderwijsleerproces. De schoolleider lijkt (te?) zeer in beslag genomen door allerlei zaken met een grotere urgentie. Het lijkt erop dat datgene wat door velen wordt gezien als bij uitstek de taak van de schoolleider in de knel komt door allerlei zaken die direct om actie of oplossingen vragen.

Al eerder is gezegd dat de taak van de schoolleider veelomvattend en divers is; zijn activiteiten zijn dan ook moeilijk te vangen onder een beperkt aantal noemers. Daarom zijn allerlei niet bij de vorige categorieën onder te brengen activiteiten noodgedwongen bijeengenomen in een rest-rubriek. Dit rest-karakter houdt echter geen oordeel in over het relatieve belang van de activiteiten binnen deze categorie; er zullen schoolleiders zijn die moeten constateren dat de hier aangeduide activiteiten naast het lesgeven de meeste tijd vragende aspecten van het schoolleiderschap in de praktijk inhouden.

Tabel 5 geeft een gevarieerd scala aan activiteiten te zien waaruit nogmaals blijkt dat de taak van de schoolleider divers en gefragmenteerd is. Vele activiteiten vragen elk op zich relatief weinig tijd, maar het totaal aantal taken is groot. Kennelijk vraagt en krijgt vrijwel alles zijn aandacht. De vraag dringt zich dan ook op of een dergelijk gedrag van de schoolleider efficiënt, effectief en dus gewenst is.

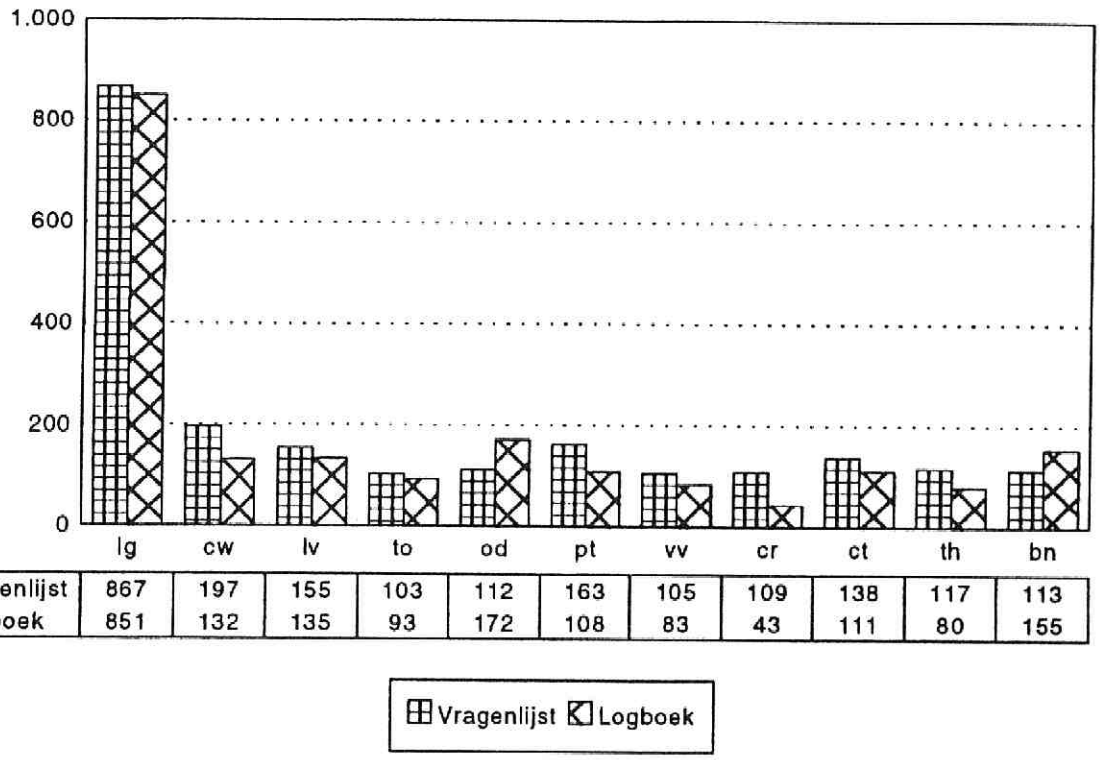
Tabel 5. Gemiddelde, door de schoolleider, bestede tijd (in minuten per week) en standaarddeviatie per onderscheiden activiteit binnen de categorie "diverse activiteiten".

Totaal bestede tijd aan activiteit	Gemiddeld aantal minuten	Standaard deviatie	Minimum	Maximum
Magazijnbeheer	11	13	0	60
Stencillen/kopiëren	41	36	0	135
Distributie melk, folders e.d.	4	10	0	45
Huisbezoek bij ouders	15	24	0	120
Uitvoeren kleine reparaties	19	26	0	150
Toezicht overblijvende leerlingen	16	39	0	203
Rooster-activiteiten	6	5	0	28
Koffie/thee zetten, afwassen etc.	13	22	0	135
Schoonmaakwerkzaamheden	17	25	0	135
Afsluiten van gebouwen	31	31	0	225
'Boodschappen' doen	20	24	0	120
Pleinwacht	36	28	0	135
(Supervisie) schoolkrant	7	10	0	46
Versieren school bij feestdagen	5	6	0	48
Beheer van speciale apparatuur	9	15	0	68
Surveilleren in de school	15	24	0	135
Toezicht onderw.onderst.personeel	20	29	0	180
Beheer schoolbibliotheek	8	13	0	60
Regelen interne verhuizingen	5	28	0	270
Volgen van nascholingscursussen	36	51	0	300
Bijhouden van het vak	77	54	3	360
Organisatie sporttoernooien	4	7	0	46
Organisatie excursies	9	23	0	192
Organisatie van toneelvoorstellingen	6	14	0	111
Overige buitenschoolse activiteiten	7	11	0	60

De schoolleider is aan diverse conciërge-taken meer dan twee uur per week "kwijt". Sommige schoolleiders zeggen hieraan nog geen minuut per week te besteden, terwijl anderen zeggen hieraan meer dan een dag per week te besteden. Onder de categorie "toezicht houden" zijn activiteiten opgenomen zoals pleinwacht, surveilleren in de school, toezicht houden op ondersteunend personeel e.d. Dit toezicht houden vraagt van de schoolleider bijna twee uur per week. Sommigen zeggen hier gemiddeld bijna een dag (van 8 uur) per week aan kwijt te zijn. Er zijn echter ook schoolleiders die hier nog geen minuut aan besteden.

Opvallend is de geringe hoeveelheid tijd die aan huisbezoek wordt besteed. Hoewel wellicht niet meer passend in deze tijd waar de persoonlijke levenssfeer van ieder gezin uitdrukkelijk wordt geëerbiedigd, is het van de andere kant denkbaar dat, gezien het grote aantal probleemleerlingen, waarvan veelvuldig sprake is, huisbezoek door de schoolleider in die gevallen juist geboden is om deze kinderen, na kennisname van de thuissituatie, op school beter te kunnen opvangen.

Aan nascholing en aan het bijhouden van het vak wordt per week bijna twee uur gespendeerd. Ook hier weer grote verschillen die uiteenlopen van 3 minuten per week tot 11 uur per week. De schoolleider besteedt ongeveer 13 uur per week aan vergaderingen, bijna 20 uur aan het verzorgen van onderwijs, 10 uur aan bureau-activiteiten, ruim 3 uur per week aan toezicht en beheer en eveneens ruim 3 uur aan diverse activiteiten waaronder bij- en nascholing en extra-curriculaire activiteiten. In totaal heeft de gemiddelde schoolleider een weektaak van ongeveer 50 uur per week. De verschillen tussen schoolleiders zijn echter groot; er zijn schoolleiders die meer dan 60 uur werken, terwijl er ook schoolleiders zijn die minder dan 40 uur werken.



	lg	cw	lv	to	od	pt	vv	cr	ct	th	bn
Vragenlijst	867	197	155	103	112	163	105	109	138	117	113
Logboek	851	132	135	93	172	108	83	43	111	80	155

lg = tijd besteed aan lesgeven
cw = tijd besteed aan correctiewerkzaamheden
lv = tijd besteed aan lesvoorbereiding
to = tijd besteed aan teamoverleg
od = tijd besteed aan overleg met individuele docenten
pt = tijd besteed aan post doornemen
vv = tijd besteed aan voorbereiden van vergaderingen
cr = tijd besteed aan correspondentie
ct = tijd besteed aan conciërge-activiteiten
th = tijd besteed aan toezicht en beheer
bn = tijd besteed aan bij- en nascholing

Fig. 1. Activiteiten van de schoolleider vergeleken voor vragenlijst en logboek. Gemiddelde per week bestede tijd in minuten.

De via de logboekprocedure verkregen gegevens leveren een beeld op dat in grote lijnen vergelijkbaar is met de gegevens verkregen via de vragenlijsten (vgl. onderstaande figuur 1). De per week bestede tijd, zoals weergegeven in de logbocken, ligt iets lager dan is geconcludeerd op basis van de via vragenlijsten verkregen gegevens. In vergelijking met zijn collega's die de vragenlijsten invulden, besteden zij meer tijd aan overleggen en vergaderen, terwijl zij op de andere terreinen zeggen minder tijd te besteden. De totale tijdsbesteding zoals geïnventariseerd via beide methodes lijkt vergelijkbaar; de verdeling van de tijd op de onderscheiden activiteitenclusters geeft accentverschillen te zien.

De taakbelasting van de schoolleider

De ervaren taakbelasting is in het onderzoek bepaald door bij iedere activiteit in de vragenlijst te vragen naar de mate waarin de schoolleider vindt dat de door hem of haar bestede tijd voldoende is geweest om de betreffende activiteit naar behoren uit te voeren. Zo vindt bijvoorbeeld 87% van de respondenten dat de tijd die besteed wordt aan teamvergaderingen ruim voldoende of

voldoende. In de bijlagen is van alle gestelde vragen aangegeven wat op dit punt het oordeel is van de respondenten.

In de volgende alinea's zal met name aandacht worden gegeven aan die activiteiten waaraan naar het oordeel van een groot deel van de schoolleiders onvoldoende, of zeer onvoldoende tijd wordt besteed. Het is uit de voorgaande paragrafen duidelijk geworden dat de schoolleider prioriteiten stelt en ook moet stellen. In deze paragraaf komt aan de orde welke activiteiten daardoor minder uit de verf komen dan de schoolleider zelf zou wensen.

Alleen die activiteiten worden genoemd die door meer dan 25% van de respondenten van het oordeel "onvoldoende" of "zeer onvoldoende" werden voorzien. De schoolleider geeft dus zelf aan ten aanzien van welke activiteiten hij of zij qua tijdsbesteding te kort is geschoten.

Tenminste één kanttekening moet hierbij worden gemaakt. Wanneer een schoolleider meldt geen enkele aandacht te hebben besteed aan een bepaalde activiteit is in een groot aantal gevallen de vraag betreffende het oordeel over het (on)voldoende zijn van deze tijd niet ingevuld. Men zou kunnen veronderstellen dat deze "missing" gegevens in feite binnen de categorie "(zeer) onvoldoende" zouden vallen wanneer deze vraag wel door de respondenten zou zijn ingevuld.

Wat betreft de activiteitencategorie "overleggen, bespreken en vergaderen" is het merendeel van de respondenten van mening dat aan de meeste overleg-situaties (er werden er 33 onderscheiden) "(ruim) voldoende" tijd wordt besteed. Binnen deze categorie zijn daarop twee uitzonderingen. Van de respondenten is meer dan 25% van mening dat "(zeer) onvoldoende tijd wordt besteed aan structureel overleg met individuele docenten (bijna 38%) en aan functioneringsgesprekken (bijna 32%).

De activiteitencategorie "onderwijs verzorgen" geeft een ander beeld te zien. Men is van mening dat men in (zeer) onvoldoende mate toekomt aan het voorbereiden van het onderwijs aan de eigen groep (bijna 38%), het verzorgen van remediële activiteiten in verband met mocilijk lerende kinderen in andere groepen dan de eigen groep (bijna 52%), het bijhouden van het vak (meer dan 70%) en het begeleiden van ouders met een onderwijstaak in de school (28%).

Ten aanzien van de 28 onderscheiden "bureau-activiteiten" is een groot aantal schoolleiders van mening dat er in (zeer) onvoldoende mate tijd aan wordt besteed. Zo is 25% van de schoolleiders van mening dat men onvoldoende tijd besteedt aan het voorbereiden van de teamvergaderingen. Van de respondenten is bijna 27% die mening toegedaan daar waar het de afhandeling van financiële zaken betreft. Ook aan het bijhouden van leerlingdossiers besteedt men onvoldoende tijd (34%). Dit geldt in nog sterkere mate voor de tijd besteed aan de vormgeving van het onderwijsleerproces: 54% van de respondenten besteedt hier naar eigen zeggen onvoldoende tijd aan. Ook komt men in onvoldoende mate toe aan het bestuderen van het toegezonden O&W-materiaal (54%). Het opstellen van activiteitenplannen tenslotte kan ook niet rekenen op voldoende tijdsinvestering van de zijde van de schoolleiders (25%).

Aan de activiteiten binnen de categorie "toezicht en beheer" (19 activiteiten) besteedt men in de meeste gevallen, naar eigen zeggen, (ruim) voldoende tijd. Een paar uitzonderingen daarop. Bijna 30% is van mening dat men onvoldoende tijd besteedt aan huisbezoek bij leerlingen. Ook besteedt men onvoldoende tijd aan het uitvoeren van kleine reparaties (33%). Het beheer van de schoolbibliotheek komt er naar de mening van 42% van de schoolleiders bekaaid af.

Taken door de schoolleider gedelegeerd aan adjunct en andere leden van het schoolteam

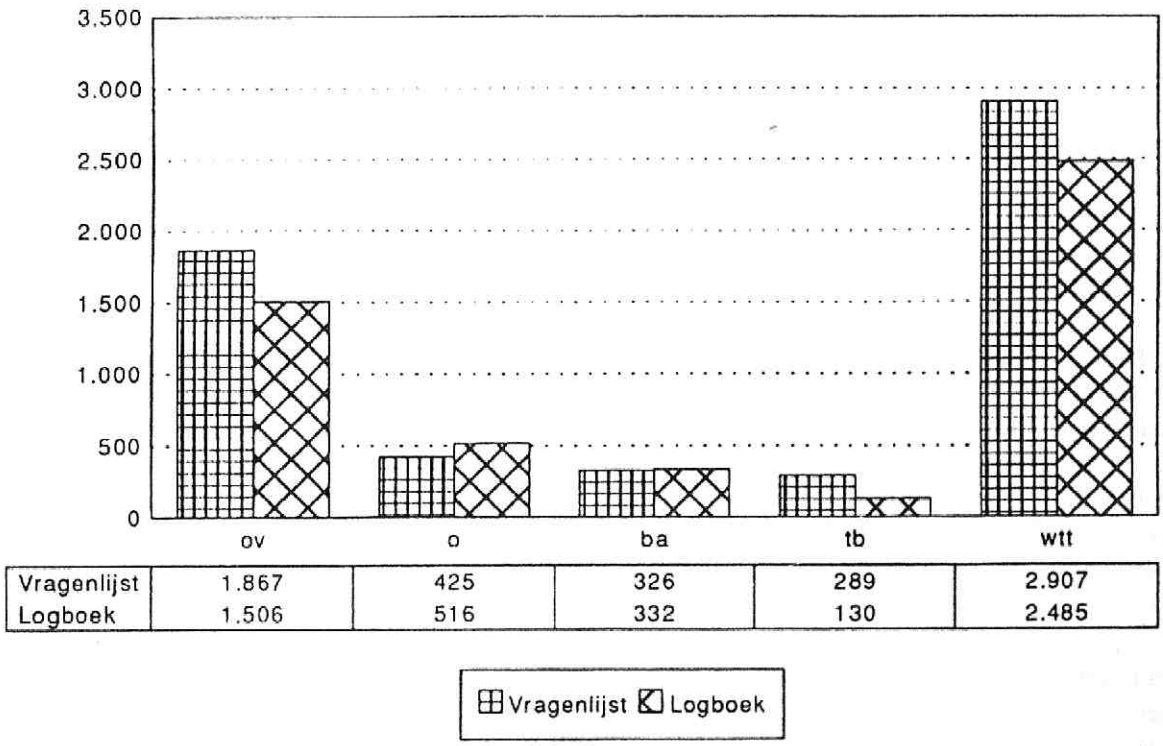
De schoolleider staat niet alleen wanneer er taken moeten worden uitgevoerd op het niveau van de school. Hij heeft de mogelijkheid een beroep te doen op de adjunct-directeur, wanneer de school althans groot genoeg is om daar de beschikking over te hebben, en hij kan ook een beroep doen op de andere collega's. In deze paragraaf wordt beschreven welke taken door de schoolleider worden gedelegeerd en hoeveel tijd door anderen dan de schoolleider aan deze taken wordt besteed.

In totaal 50 adjunct-schoolleiders retourneerden een ingevulde vragenlijst en 46 adjuncten zonden een bijgehouden logboek terug. Deze aantallen zijn iets lager dan het geval was bij de schoolleiders; voor het grootste deel valt dit te verklaren uit het feit dat niet iedere school groot genoeg is om een adjunct te hebben.

De totale omvang van de weektaak van de adjuncten die een vragenlijst invulden verschilt duidelijk van die van de adjuncten die het logboek bijhielden. Daar waar deze laatsten uitkomen op een weektaak van ongeveer 40 uur zeggen hun collega's die de vragenlijst invulden uit te komen op een weektaak van bijna 50 uur, vergelijkbaar dus met de tijd die de schoolleider aan zijn taken zegt te besteden.

Er zijn verschillen te constateren tussen de resultaten verkregen via het logboek respectievelijk via de vragenlijst; deze verschillen zijn voor een groot deel terug te voeren op de verschillen in omvang van de activiteiten die samenhangen met het lesgeven tussen beide groepen; de adjuncten die de vragenlijst retourneerden, besteedden zes uur per week meer aan lesgeven dan hun collega's die de logboeken bijhielden. Dit verschil kan, in ieder geval ten dele, verklaard worden uit het "dubbel declareren" van bijvoorbeeld correctiewerkzaamheden die "onder de les" zijn uitgevoerd, waar al eerder over is gesproken. Een ander deel kan wellicht verklaard worden uit het feit dat de "vragenlijst"-adjuncten aan scholen zijn verbonden die kleiner zijn dan de scholen van de "logboek"-adjuncten; de hoeveelheid ambulante tijd die voor beide groepen beschikbaar is zal dus kunnen verschillen en daarmee kan worden verklaard dat de "logboek"-adjuncten minder tijd aan het verzorgen van onderwijs besteden.

Uit figuur 2 blijkt dat adjuncten naast hun reguliere onderwijstaken ruim 16 uur (logboek) respectievelijk ruim 17 uur (vragenlijst) besteden aan taken die geen verband houden met de eigen onderwijsgroep. De adjuncten die het logboek bijhielden overleggen en vergaderen ruim 8 uur per week; de collega's die de vragenlijst invulden kwamen op net 7 uur per week uit. Aan de



	ov	o	ba	tb	wtt
Vragenlijst	1.867	425	326	289	2.907
Logboek	1.506	516	332	130	2.485

ov = tijd besteed aan onderwijs verzorgen
o = tijd besteed aan overleggen en vergaderen
ba = tijd besteed aan bureauactiviteiten
tb = tijd besteed aan toezicht en beheer
wtt = tijd besteed aan een totale weektaak

Fig. 1. Activiteiten van adjunct-schoolleider vergeleken voor vragenlijst en logboek. Gemiddelde per week bestede tijd in minuten.

zgn. bureau-activiteiten besteden beide categorieën ongeveer 5 uur per week. Aan activiteiten ondergebracht in de categorie "toezicht en beheer" zeggen de vragenlijst-adjuncten daaraan eveneens 5 uur per week te besteden, terwijl de logboek-adjuncten op iets meer dan de helft daarvan uitkomen.

Evenals het geval was bij de gegevens van de schoolleiders is de spreiding in de resultaten bijzonder groot te noemen; de verschillen in gemiddelde tijdsbesteding van adjuncten zijn aanzienlijk.

De gemiddelde adjunct blijkt evenals de directeur een weektaak te hebben van ongeveer 50 uur. Daarbij blijken adjuncten dezelfde taken te verrichten als de directeuren. Op individuele scholen kan (en zal) uiteraard vaak sprake zijn van een taakverdeling; wanneer men de groep adjuncten echter als geheel bekijkt, blijken er geen taken te zijn die echt duidelijk tot de taak van de directeur of tot die van de adjunct worden gerekend. Aangezien adjuncten een onderwijstaak hebben die, op de meeste scholen althans, groter is dan die van de directeur, besteden adjuncten wel minder tijd dan de schoolleiders aan de diverse activiteiten die men op het niveau van de school kan onderscheiden. Er blijkt, voor de groep als geheel, dus geen duidelijk verschil te bestaan in de inhoud van het takenpakket van schoolleider en adjunct.

Zowel via de vragenlijsten als via de logboeken is de respondent-schoolleiders expliciet gevraagd uitvoerig aan te willen geven welke taken die een onderwijsgroepoverstijgend karakter hebben, gedelegeerd worden aan andere teamleden dan de (eventueel aanwezige) adjunct-schoolleider. Niet alleen is gevraagd naar de aard van deze taken maar ook om een schatting te geven van de hoeveelheid tijd die daar, per onderscheiden en zelf te kiezen, tijdseenheid mee is gemoeid.

Een groot aantal verschillende taken blijkt door schoolleiders te worden gedelegeerd aan teamleden. De nadruk ligt daarbij op "toezicht en beheerstaken" en op "extra-curriculaire zaken". Ook worden relatief veel taken die eerder tot de bureau-activiteiten van de schoolleider werden gerekend aan teamleden gedelegeerd. Alleen op het punt van de diverse overlegsituaties die men op een school kan onderscheiden wordt weinig door de schoolleider gedelegeerd. Tot slot moet nog genoemd worden dat het verzorgen van remediële activiteiten veelvuldig wordt gedelegeerd en qua omvang van de taak de meeste tijd van andere teamleden vergt.

Opnieuw verkrijgt men een beeld van een breed scala aan activiteiten dat gedelegeerd wordt. De gemiddelde totale tijdsomvang van de gedelegeerde taken per teamlid is gering te noemen, nl. ongeveer 8 à 9 uur per week voor de groep (overige) teamleden als geheel.

Omvang en inhoud van de taken op schoolniveau die aan andere teamleden worden gedelegeerd

Zowel via de vragenlijsten als via de logboeken is de respondent-schoolleiders expliciet gevraagd uitvoerig aan te willen geven welke taken die een onderwijsgroepoverstijgend karakter hebben, gedelegeerd worden aan andere teamleden dan de (eventueel aanwezige) adjunct-schoolleider. Niet alleen is gevraagd naar de aard van deze taken maar ook om een schatting te geven van de hoeveelheid tijd die daar, per onderscheiden en zelf te kiezen, tijdseenheid mee is gemoeid. De antwoorden zijn gescoord op dezelfde manier als dat bij de scoring van de logboeken is geschied. Bij de berekeningen zijn de tijden per gekozen frequentie omgerekend naar een tijdsbesteding in minuten per week

Een groot aantal verschillende taken blijkt door schoolleiders te worden gedelegeerd aan teamleden. De nadruk ligt daarbij op "toezicht en beheerstaken" en op "extra-curriculaire zaken". Ook worden relatief veel taken die eerder tot de bureau-activiteiten van de schoolleider werden gerekend aan teamleden gedelegeerd. Alleen op het punt van de diverse overlegsituaties die men op een school kan onderscheiden wordt weinig door de schoolleider gedelegeerd. Tot slot moet nog genoemd worden dat het verzorgen van remediële activiteiten veelvuldig wordt gedelegeerd en qua omvang van de taak de meeste tijd van andere teamleden vergt. En opnieuw blijken de uitkomsten verkregen van de logboeken vergelijkbaar met die welke via de vragenlijsten zijn verkregen.

Wanneer men het overzicht van de gedelegeerde taken niet beschouwt met de vraag *welke*

taken worden gedelegeerd maar veeleer met de vraag hoeveel *tijd* er is gemoeid met de uitvoering van deze gedelegeerde taken, dan valt op dat dit bijzonder weinig is. Een toelichting lijkt op zijn plaats. Een gemiddeld schoolteam bestaat (in dit onderzoek) uit 11 personen, die niet altijd allen fulltime zullen werken (het gemiddeld aantal leerlingen op de scholen in het onderzoek was 171 respectievelijk 181). Afgezien van directeur en adjunct-directeur werken er op de gemiddelde school dus nog 9 leerkrachten. De school zal dus al gauw zes tot zeven groepen hebben die elk een (fulltime) groepsleerkracht nodig hebben (of meer deeltijders). Uitgaande van zeven groepen en een fulltime aangestelde directeur en adjunct-directeur zijn er dus, naar fulltime-equivalenten omgerekend, tenminste vijf groepsleerkrachten beschikbaar op de gemiddelde school voor het uitvoeren van gedelegeerde taken op schoolniveau. Op die fictieve gemiddelde school blijken er in zijn totaliteit voor 9 à 10 uur per week taken die op leergroepoverstijgend niveau liggen te worden gedelegeerd aan deze vijf groepsleerkrachten; gemiddeld dus nog geen twee uur per week per groepsleerkracht.

Samenhangen tussen achtergrondkenmerken van scholen en de tijdsbesteding van schoolleiders

Zowel uit de analyses van de gegevens afkomstig van de vragenlijsten als van de gegevens, die afkomstig zijn van de logboeken, blijkt dat er nauwelijks samenhang tussen achtergrondgegevens en tijdsbesteding per activiteit kon worden aangetoond. Zo kon geen enkele samenhang worden aangetoond tussen het gemiddeld aantal jaren ervaring van de schoolleider als schoolleider, diens gemiddelde aantal jaren ervaring in het basisonderwijs, de gemiddelde leeftijd van het schoolteam, de denominatie van de school, de sekse van de schoolleider enerzijds en de tijd die besteed wordt aan de diverse onderscheiden activiteiten anderzijds. Een aantal uitzonderingen daarop zijn wel aan te geven. In de eerste plaats blijkt de vrouwelijke schoolleider aanzienlijk minder tijd aan onderwijsgevende werkzaamheden voor de eigen groep te besteden dan haar mannelijke collega's. Vrouwelijke schoolleiders blijken in dit onderzoek dan ook schoolleider te zijn van gemiddeld grotere scholen dan van hun mannelijke collega's (gemiddeld 218 leerlingen versus 191 leerlingen). Zij hebben daarmee dus de beschikking over meer ambulante tijden en zullen daarom dus minder tijd besteden aan onderwijsgevende taken.

Vervolgens blijkt dat op scholen waar een conciërge aanwezig is door de schoolleider meer tijd wordt besteed aan overlegsituaties en minder tijd aan onderwijs verzorgen. Ook deze relaties zijn terug te voeren op het verschil in schoolgrootte van scholen met en zonder conciërge (233 resp. 163 leerlingen).

In feite kan er geen enkele relatie worden aangetoond tussen achtergrondkenmerken van de school, de schoolleider en het schoolteam die niet direct een logisch gevolg zijn van de schoolgrootte. Vervolgens zijn de wel geconstateerde relaties onbeduidend: op grotere scholen leidt dit tot meer overleg en minder activiteiten die samenhangen met het lesgeven van de schoolleider dan op kleinere scholen het geval is.

Betekenisvoller zijn echter de relaties die *niet* aangetoond konden worden en die men wel zou mogen verwachten. De schoolleider van de grotere school (met conciërge) besteedt kennelijk niet méér tijd aan het nadenken over een schoolontwikkelingsbeleid, het opstellen van activiteitenplannen, het houden van functioneringsgesprekken, het volgen van bij- en nascholingscursussen, het bijhouden van leerlingdossiers, etc. en tevens niet minder tijd aan zaken als stencillen en kopiëren, pleinwacht of het regelen van financiële zaken, etc. dan de schoolleider van een kleinere school (zonder conciërge). Er zijn dus geen aanwijzingen dat schoolleiders van grotere scholen een ander tijdsbestedingspatroon hebben dan schoolleiders van kleinere scholen.

Evenals het geval was bij de berekening van de samenhang tussen de tijdsbesteding per activiteit van schoolleiders en achtergrondkenmerken van school, schoolleider en schoolteam worden opnieuw nauwelijks relaties geconstateerd wanneer men deze achtergrondkenmerken in verband brengt met de tijdsbesteding van adjunct-schoolleiders of de tijdsbesteding van het team als geheel aan taken op het niveau van de school. Schoolgrootte, ervaring, leeftijd, denominatie, percentage allochtone leerlingen blijken nauwelijks of niet samen te hangen met de

tijdsbesteding van een schoolteam. De aannames achter de schaal- en autonomievergrotingsoperatie gaan er vanuit dat daardoor de mogelijkheden van scholen en schoolteams om beleid te maken toenemen; de situatie in de onderwijspraktijk wijst uit dat men nauwelijks profiteert van deze, in theorie bestaande, mogelijkheden.

CONCLUSIES OP HOOFDPUNTEN EN AANBEVELINGEN

Als de meest opvallende conclusie komt naar voren dat schoolleiders nauwelijks toekomen aan datgene wat velen zien als hun primaire taak, namelijk de vormgeving van het onderwijsinhoudelijk leiderschap van de school. Zowel de verzamelde gegevens afkomstig van vragenlijsten als van de logboeken leveren dit beeld op. Men komt kennelijk als gevolg van allerlei zaken die, op de korte termijn, een hogere prioriteit hebben niet toe aan deze taak van de schoolleider. In paragraaf 2 is betoogd dat de visie waarin de schoolleider in belangrijke mate wordt gezien als onderwijskundig leider één der vele rollen is die de schoolleider in de loop der tijd en in verschillende landen is en wordt toebedeeld. De aandacht die in Nederland bestaat voor effectieve scholen en schoolleiders (vgl. Scheerens, 1989) en effectieve instructie (Creemers, 1991) is een indicatie van het belang dat in dit land wordt gehecht aan juist deze rol van de schoolleider. Maar het moet beklemtoond worden dat dit één der vele aan de schoolleider toegedachte rollen is. Aan juist deze rol van onderwijskundig leider komt hij in slechts beperkte mate toe. Vele van zijn andere rollen vragen veel meer van zijn tijd. Of dit een (on)gewenste situatie is, valt op basis van dit onderzoek niet te bepalen.

Een tweede belangrijke conclusie die getrokken kan worden is dat de taakomvang, maar vooral de tijdsbesteding per activiteit, sterk verschilt van schoolleider tot schoolleider. Achtergrondkenmerken van de school, de schoolleider of het team als geheel lijken geen samenhang te vertonen met de tijd die aan de diverse activiteiten wordt besteed. Ook de ondernomen pogingen om enige samenhang te vinden in de aan de diverse activiteiten bestede tijd hebben geen resultaat gehad. Men kan zich derhalve moeilijk aan de indruk onttrekken dat er weinig lijn zit in de manier waarop de schoolleiders basisonderwijs, als groep, invulling geven aan hun taken.

Een volgende gevolgtrekking betreft het (niet-) delegeren van taken door schoolleiders. Schoolleiders blijken in beperkte mate taken, onder behoud van eindverantwoordelijkheid uiteraard, te delegeren aan andere teamleden. Ook op dit punt bestaan grote verschillen tussen scholen; de trend is echter dat het delegeren van taken (te) weinig gebeurt. Ongeacht de oorzaak hiervan, ligt hier een duidelijke mogelijkheid om de schoolleider te ontlasten.

Als vierde in 't oog springende opbrengst van het onderzoek kan genoemd worden de constatering dat (zeer) veel tijd van de schoolleider en van de overige teamleden wordt besteed aan (de voorbereiding van) vergaderingen en andere overlegsituaties. Wanneer men erin zou slagen deze tijd (aanzienlijk) terug te brengen zou op dit punt grote winst te boeken zijn.

Een vijfde conclusie betreft het takenpakket van de adjunct-schoolleider. Uit het onderzoek blijkt dat adjuncten, gemiddeld genomen, *niet* andere taken uitvoeren dan schoolleiders. Vanwege het minder ambulant zijn van adjuncten, besteden zij wel minder tijd aan deze taken dan de schoolleider. Het onderzoek wijst duidelijk uit dat er geen consensus bestaat over hetgeen een taak "des schoolleiders" is en wat een taak "des adjuncts" is.

Vervolgens kan als zesde opmerkelijke uitkomst worden genoemd dat samenhang tussen enerzijds achtergrondkenmerken van de school, de schoolleider en het schoolteam en anderzijds de tijd die besteed wordt aan de diverse onderscheiden activiteiten van de schoolleider, van zijn adjunct of van het team als geheel, vrijwel ontbreekt. Ook blijkt er geen relatie te kunnen worden aangetoond tussen de ervaren taakbelasting van de schoolleider en de tijd die hij daadwerkelijk besteedt aan de diverse activiteiten.

In de volgende alinea's wordt ingegaan op de mogelijkheden om de realisatie van de doelstellingen van het onderwijsbeleid dichterbij te brengen door een aantal aanbevelingen te formuleren.

In de eerste plaats lijkt het aanbevelenswaardig om in de primaire opleiding veel meer dan nu

het geval is, aandacht te besteden aan de concrete werkzaamheden van alledag waar een schoolleider mee te maken heeft. Niet alleen potentiële schoolleiders maar alle aankomende leraren basisonderwijs zouden, voordat zij als leraar aan het werk gaan, moeten weten wat er komt kijken bij het runnen van een school. Door middel van praktijksimulaties zouden vaardigheden op dit terrein kunnen worden aangeleerd. Het is belangrijk dat alle leraren van een school weten wat er zich op leergroepoverstijgend niveau afspeelt op een school opdat een schoolleider met succes een beroep kan doen op deze leraren wanneer hij taken wil en moet delegeren. Het aanleren van theoretische kennis en het opdoen van praktische ervaring met zaken zoals: het voeren van een schooladministratie, de functie en taak van het schoolbestuur, het opzetten, bijhouden en raadplegen van leerlingvolgsystemen, het beheren van een schoolbibliotheek, schoolarchief en magazijn, alsmede het gebruik en de toepassing van diverse onderwijsondersteunende apparatuur, (inclusief de computer in zowel de functie van administratief- als van onderwijskundig hulpmiddel) zou een integraal onderdeel moeten zijn van de initiële opleiding tot leraar basisonderwijs. Het management van basisscholen kan en mag dus niet het exclusieve domein zijn van de schoolleider en van eventueel zijn adjunct. Wanneer een schoolleider ervan kan uitgaan dat deze elementaire kennis en basisvaardigheden aanwezig zijn bij de docenten op zijn school zal hij gemakkelijker en met meer succes een beroep kunnen doen op zijn collega-docenten.

De voorgaande aanbeveling houdt ook verband met een uitkomst van het onderzoek waaruit blijkt dat de gemiddelde schoolleider weinig taken of onderdelen daarvan delegeert (of denkt te kunnen delegeren) aan andere leden van het schoolteam. Wanneer een schoolleider meer ruimte wil krijgen voor zijn primaire taken is het noodzakelijk dat hij, meer dan nu het geval is, taken delegeert (en taken kan delegeren) aan andere teamleden. Daartoe is het noodzakelijk dat bij sommige schoolleiders en bij sommige schoolteams een omslag in het denken plaatsvindt. Het runnen van de school moet door het gehele team als een collectieve verantwoordelijkheid worden gezien. Dit betekent dat niet de schoolleider "alles" moet doen (omdat hij daarvoor betaald zou worden) maar dat hij eindverantwoordelijke is en dat hij leiding geeft aan het geheel. De resultaten van het onderzoek geven aanleiding tot de veronderstelling dat er nogal wat schoolteams bestaan waar de schoolleider, en eventueel zijn adjunct, werkweken realiseren van 50 uur of meer, terwijl tegelijkertijd de overige teamleden niet veel verder komen dan 30 à 35 uur per week. Dit is een onrechtvaardige verdeling van de taakbelasting op een school. Deze scheve verdeling wijst wellicht ook op een ongewenste perceptie van de taak van de schoolleider door zowel het schoolteam als vaak door hemzelf. Er zijn schoolleiders die taken willen delegeren maar een onwillig team tegenover zich vinden; er zijn echter ook schoolleiders die om welke reden dan ook niet in staat blijken om taken uit handen te geven. In beide situaties moet naar een oplossing worden gezocht, zodanig dat meer taken worden gedelegeerd.

Uit het onderzoek blijkt dat de gemiddelde schoolleider relatief veel tijd kwijt is aan wat kortheidshalve wordt omschreven als conciërge-activiteiten. Hoewel veel schoolleiders deze activiteiten wellicht voor hun eigen gevoel noodgedwongen uitvoeren en mede als gevolg daarvan niet voldoende toekomen aan hun primaire onderwijskundige taken, zijn dit eigenlijk activiteiten die niet (zouden moeten) behoren tot hun takenpakket. Ook op scholen waar een conciërge aanwezig is besteden schoolleiders hier een substantieel deel van hun tijd aan. Het verdient aanbeveling dat schoolleiders ophouden dergelijke taken uit te voeren. Dit kan, door deze taken te delegeren aan een conciërge, aan andere teamleden of, als een school zich geen conciërge kan veroorloven, aan een persoon die met een dergelijk takenpakket door meerdere scholen tegelijk met dat doel in dienst is genomen. Het is weinig professioneel wanneer een gemiddelde schoolleider meer dan twee uur van zijn kostbare tijd per week letterlijk "kwijt" is aan zaken als stencillen en kopiëren, het uitvoeren van kleine reparaties, koffie- en theezetten, afwassen, magazijn en apparatuurbeheer e.d.

Veel tijd van schoolleiders wordt doorgebracht in overlegsituaties: teamoverleg, overleg met ouders, met het bevoegd gezag, met leerkrachten, etc. Van de totale tijd die de schoolleider buiten zijn onderwijstaak aan de school besteedt, wordt 35% doorgebracht in overlegsituaties. Wanneer men de tijd die wordt besteed aan de voorbereiding daarvan daar nog eens bij optelt

dan blijkt, dat gemiddeld genomen, bijna de helft van de beschikbare tijd die de schoolleider voor de school beschikbaar heeft aan overleggen en vergaderen wordt besteed. Het is duidelijk dat er veel "winst" is te boeken wanneer de tijd die wordt besteed aan overleggen en vergaderen wordt teruggebracht. Vooral het opheffen of beperken van het veelal wekelijkse team-overleg (meer dan 90% van de scholen heeft wekelijks teamoverleg) zou veel ruimte kunnen scheppen voor niet alleen de schoolleider zelf maar voor het team als geheel. Wanneer er per week, gemiddeld anderhalf uur, door het team als geheel wordt vergaderd (dit blijkt uit het onderzoek) dan vraagt dat tenminste 14 uur per week van het gemiddelde schoolteam (in het geval van 8 aan het teamoverleg participerende personen) als geheel. Een tweewekelijks of driewekelijks overleg met het team als geheel verdient, zo lijkt het, de voorkeur boven de thans geldende situatie. Wanneer een schoolleider afspraken wil maken met een individuele leraar dan behoeft hij uiteraard niet te wachten tot het volgende teamoverleg, maar kan hij direct "zaken" doen met die leraar. Wanneer tegelijkertijd meer gedelegeerd wordt kan het teamoverleg met name besteed worden aan voortgangsrapportages en terugkoppeling. Hoewel een dergelijke handelwijze op individuele scholen wel zal plaatsvinden, wordt deze voor de hand liggende en tijdbesparende strategie op de meeste scholen (nog) niet toegepast.

Omdat uit het onderzoek blijkt dat er nauwelijks een lijn valt te ontdekken in het taak- en tijdsbestedingspatroon van schoolleiders verdient het wellicht aanbeveling om op basis van deze onderzoeksresultaten, die een toch wel grote discrepantie tussen de doelstellingen van het onderwijsbeleid en de gerealiseerde praktijk aan het licht brengen, realistische beroepsprofielen van schoolleiders basisonderwijs te ontwikkelen waarbij niet het uitgangspunt is wat deze schoolleider allemaal zou moeten doen maar wat hij allemaal zou kunnen doen gezien zijn huidige taakinvulling. Bij de ontwikkeling van dergelijke profielen zou tegelijkertijd expliciet moeten worden betrokken welke taken tot die van de schoolleider behoren en welke taken andere personen en instanties zoals schoolbestuur, administratiekantoor en schoolbegeleidingsdienst voor hun rekening zouden moeten nemen. En tenslotte moet daarbij worden betrokken welke taken de schoolleider met recht en reden kan delegeren aan andere teamleden. Het lijkt noodzakelijk dat op korte termijn over deze taakinvulling van de schoolleider consensus gaat ontstaan. De huidige situatie wordt op dit punt gekenmerkt door een verregaande onduidelijkheid; het lijkt er op dat op iedere school een ander evenwicht wordt gezocht; veel schoolleiders voelen zich in de steek gelaten omdat alle betrokkenen: de ouders, de collega-docenten, het schoolbestuur, de overheid niet te vergeten en begeleidende en nascholende instanties hem vooral komen vertellen wat hij allemaal zou moeten doen, maar dat er weinigen zijn die hem ontlasten van oneigenlijke taken, opdat hij deze verwachtingen ook inderdaad kan waarmaken. Bovendien blijken de meningen van alle betrokkenen op dit punt zeer verdeeld. Op grond van de onderzoeksresultaten is geen eenduidig advies te geven over de verdeling van taken die op een school(leider) afkomen. De concrete context waarin een school(leider) functioneert moet wellicht bepalend zijn voor de mate waarin hij taken delegeert en aan wie hij deze taken delegeert.

LITERATUUR

Beck, L.G. & Murphy, J. (1993). *Understanding the Principalship*. New York: Teachers College, Columbia University.

Brookover, W.B., Beady, C., Flood, P., Schweitzer, J. & Wisenbaker, J. (1979). *School Social Systems and Student Achievements: Schools can make a difference*. New York: Praeger Publishers.

Centraal Bureau voor de Statistiek (1991). *Zakboek onderwijsstatistieken*. Den Haag: CBS.

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

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

Emans, B. & Veen, P. (1983). Macht en machtsuitoefening op scholen. *Handboek Schoolorganisatie*. 2210-2222.

Gielen, K. (1983). Wat doen schoolleiders eigenlijk? A: contacten, werkkarakteristiek en rolgedrag B: beslissen naar eigen inzicht. *Meso, 14-15*; 19-25; 20-24.

Gielen, K. & Kraaykamp, J. (1981). Een analyse van de schoolleidersfunctie. *Meso, 2,* 2-7.
Giesbers, J. & Marx, E. (1990). Topmanagement en schoolorganisatie. *Meso, 52,* 12-15.
Grift, W. van de, (1987). *Zelfpercepties van onderwijskundig leiderschap en gemiddelde leerlingprestaties.* Paper ORD, 26-27 mei 1987 te Groningen. Amsterdam: SCO.
Hallinger, P. & Murphy, J. (1987). *Social context effects on school effects.* Paper AERA, Washington.
Kmetz, J.T. & Willower, D.J. (1982). Elementary-school principals work behavior. *Educational administration quarterly, 18* (4), 62-78.
Levine, D.H. & Lezotte, L.W. (1990). *Unusally effective schools.* National center for effective schools research and development.
Martin, W.J. & Willower, D.J. (1981). The managerial behavior of high school principals. *Educational administration quarterly, 17,* 69-90.
Morris, V.C. (1982). The urban principal. *Phi Delta Kappan,* 689-692.
Rutter, M., Maughan, B. Mortimore, P. & Ouston, J. (1979). *Fifteen thousand hours. Secondary schools and their effects on children.* London: Open Books.
Scheerens, J. (1990). Conceptualisering van schoolmanagement. *Onderwijskundig Lexicon,* Editie II, 1-35. Alphen aan de Rijn: Samsom.
Stoel, W.G.R. (1994). *De taakinhoud, de taakomvang en de taakbelasting van schoolleiders in het basisonderwijs.* Enschede: Onderzoekscentrum Toegepaste Onderwijskunde (OCTO).
Wellisch, J.B., MacQueen, A., Carriere R., & Duck, G. (1978). Management and organization in successful schools. *Sociology of Education, 51,* 211-226.

Manuscript ontvangen 28-4-1994
Definitieve versie ontvangen 2-9-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 2, pp. 154-164

A self-evaluation procedure for schools using multilevel modelling

Roel J. Bosker and Jaap Scheerens
University of Twente, the Netherlands

ABSTRACT

For various reasons self-evaluation is becoming more and more important in schools. After highlighting the most common approaches to self-evaluation, it is argued that self-evaluation should primarily be based on the outcomes of educational practice. When designing a self-evaluation system one has to cope with two problems: how to define a fair school effect measure, and how to locate those practices that may lead to malfunctioning. It is shown how pupil monitoring systems can be used to construct a school monitoring system. The statistical aspects of such a system can be handled well by using multilevel statistical models. The proposed approach is illustrated using data on the development of pupils in mathematics achievement. The indicator proposed for self-evaluation purposes is compared with 7 other indicators. Striking differences between indicators are found and discussed. Moreover, it is discussed how such a monitoring system could be modified to detect educational practices that lead to malfunctioning of pupils. Finally, the practical aspects of implementing the proposed system are depicted.

INTRODUCTION

One of the consequences of the Dutch national educational decentralization policy is the increased attention paid to quality control at the school level. With this increasing importance of the schools' autonomy, systems are designed and developed (and sometimes implemented) to help schools in their self-evaluation task. These include pupil-testing, management information systems, instruments for school diagnosis, etc.

In this article the attention will be focused on the application of multi-level models for the self-evaluation of the effectiveness of a school. The basis for this self-evaluation is the educational output, which is seen as the core criterion for educational quality. From this point of view there are two central problems in assessing a schools' quality:

1. how should we define a fair school effect measure and
2. how can we find those grades/teachers within a school that mostly affect the effectiveness of a particular school?

It will be demonstrated how pupil monitoring systems, that are being applied in about 35% of Dutch primary schools, can be used for the purpose of school self-evaluation. First, existing different approaches to school self-evaluation will be presented and discussed. Next the logic of using pupil monitoring systems will be described, and the statistical model to construct a school monitoring system will be introduced. This idea will be illustrated by using longitudinal data on mathematics achievement in lower general secondary schools (mavo). Eight different school effect measures will be defined, estimated and compared with each other. Furthermore it is discussed how to detect educational practices that may lead to malfunctioning. Moreover practical prerequisites of a self-evaluation procedure as proclaimed in this article will be mentioned.

* Adres: Faculteit Toegepaste Onderwijskunde, Universiteit Twente, Postbus 217, 7500 AE Enschede.
Dit is het laatste artikel behorende bij het themanummer.

APPROACHES TO SCHOOL SELF-EVALUATION

Evaluation is at the core of the concept of school effectiveness. The very question of judging the degree to which schools attain their primary objectives cannot be answered unless evaluative data, especially output data, are available. School improvement too – though this does not appear from the early improvement literature – builds heavily on the presumption that information on the current state of functioning of a school is available. In the more recent school improvement literature, to be associated with the OECD/CERI International School Improvement Project the recognition of the importance of evaluation data is clearly evident from the attention that is given to School Based Review (Hopkins, 1987). In more abstract terms, evaluation can be seen as the basic prerequisite for all types of control (management, planning, policy-making) (cf. De Leeuw, 1988).

There are three specific ways in which evaluation can be geared to the concept of school effectiveness:

1. (as defined above) evaluation of goal attainment to judge whether a school can be considered effective or not (usually this is done by comparing schools with each other);
2. the degree to which schools internally make proper use of evaluations (both at school and classroom level), which has been shown to be a favourable condition to school effectiveness; here evaluation, or "the evaluative potential" of the school is an explanatory variable with respect to the criterion mentioned above (under 1);
3. knowledge about the "process correlates" of educational achievement as established in school effectiveness research can be used in guiding the selection of relevant variables for educational evaluation systems, such as indicator systems (cf. Scheerens, 1990).

Currently, several approaches to school self-evaluation are being used. Each has a specific disciplinary background and a specific context in which the approach was originally employed, as is shown in Table 1.

Each of these approaches will be sketched briefly and strong and weak points will be discussed.

Table 1. Different origins of school self-evaluation.

approach	disciplinary background	context
school-based review	social psychology, education	schools
management information systems	business administration, operations research	private industry
educational indicators	economics, educational statistics	macro-level applications
organizational diagnosis	management consultancy	private industry, public-sector organizations
pupil monitoring systems	educational measurement	(remedial) teaching

School-based review

School-based review depends heavily on opinions of school personnel on discrepancies between the actual and an ideal state of affairs in schools. In this way a broad perspective, in which all the main aspects of school functioning can be scrutinized, is possible. Usually, respondents are also asked to indicate whether a certain discrepancy should be actively resolved. In this way this approach to school self-evaluation seeks to gear improvement-oriented action to appraisal. The context of application usually is school improvement, which means that a school-based review is carried out when there is already a certain commitment to educational innovation.

Advantages of this approach are: a broad scope, a user-friendly technology, an explicit linkage between evaluation and action, and a high degree of participation (all school personnel takes part in the review). A definite weakness of school-based review is its dependence on subjective opinions and its (usual) negligence of "hard" factual data on school functioning, most notably output data.

Examples of procedures for school-based review are the GRID and GILS-systems (see Hopkins, 1987, and the SAS-system (Voogt, 1989).

School management information systems

School management information systems have been inspired by similar systems in private industry. Generally they consist of a careful modelling of information streams and information needs within a company, deciding which data should be available for purpose on a more or less permanent basis, followed by design and implementation of a computer configuration and software. Bluhm and Visscher (1990) describe a management information system as an information system based on one or several computers, consisting of a data-bank and one or several computer applications, which enable computer-based data storage, data analysis and data distribution.

A question that could be answered by means of such a school management information system would be: "to which degree has absenteeism decreased after the implementation of specific measures to fight absenteeism?"

Management information systems have a great potential for supplying important information on a routine basis. At present there are still a lot of practical barriers: one needs to have sufficient and adequate computer hardware and even when professionally developed software packages become available, quite a few specific maintenance functions must be carried out, while new routines and perhaps even functions to guarantee adequate data-entry should be developed.

Educational indicators

Although educational indicator systems are usually employed at the macro level (the level of national educational systems), for instance to describe the "state of education" of a country on a yearly basis, some authors have suggested applications at the school level (Teauber, 1987; Oakes, 1987; Scheerens, 1990). When applied at the school level, educational indicator systems typically will include "process" or "throughput" information, next to input, school-context and output data.

Results of schooleffectiveness research studies are usually employed to select process indicators. The general idea of indicators is to provide an at-a-glance profile of certain important characteristics of an educational system.

This means that there is no aspiration to "dig deep", while employing easily measured characteristics and so-called proxy measures. This feature is at the same time a definite limitation of the approach. Another "danger" is the use of process or throughput data as evaluation criteria, instead of explanatory conditions of educational outputs. This could easily lead to goal displacement, where the "means" in education are treated as "goals" in themselves. A technical limitation which might encourage this improper use of process indicators is the fact that the question of relating process and output indicators by means of formal statistical analysis has hardly been tackled for applied purposes. This problem will be addressed in other sections of this article.

Organizational diagnosis

As educational institutes (schools and universities) are made to function more autonomously, it is quite likely that they become more like private companies in their managerial and organizational characteristics. An example of this would be a stronger emphasis on strategic planning and on scanning the external environment of the school. It is therefore not surprising that approaches used in management consultancy are introduced in schools. Although these approaches, generally labelled as "organizational diagnosis" or "management audit", usually depend on an external organizational consultant – they are also available for school self-diagnosis. In contrast to school-based review these approaches tend to be exclusively based on information provided by the management of the organization. So, when they are used without an external consultant they would appear to be somewhat like "management introspection". A strong point of this approach is that it is likely to pay attention to issues that were kept largely unnoticed by the educational province, such as external contacts, anticipation of developments in the relevant environment, and flexibility in offering new types of services. The most important disadvantage remains, however, that this approach is not so easy to transform to a school-based application, without an external consultant.

Pupil monitoring systems

The focus of attention in this article is self-evaluation at the school level. Pupil monitoring systems operate at the micro level (class level) of educational systems. In the ensuing sections of this article it will be shown how this class of techniques can also be used for self-evaluation at the school level.

Basically pupil monitoring systems are sets of educational achievement tests that are used for purposes of formative didactic evaluation. An important function is to detect which pupils fall behind and where they experience difficulties.

Pupil monitoring systems have one assett which, in our opinion, is essential for all efforts to make school functioning more effective: the centrality of output data at the level of the individual pupils measured by means of achievement tests. If approaches to school self-evaluation neglect this type of data there is a risk that the information basis they supply for educational or administrative decision-making is faulty (see the earlier reference to the phenomenon of goal displacement).

SCHOOL MONITORING SYSTEMS

In the preceding section it was argued that the evaluation of the performance of a school should be primarily based on pupil progress in achievement. With primary schools being commissioned to refer less pupils to special education than before, schools try to assess pupil progress in order to detect, diagnose and remedy learning problems. For this purpose they use achievement tests. The Dutch National Testing Service provides tests, to be administered twice during the school year (Gillijns & Verhoeven, 1991). These tests meet the standards of Item Response Theory (IRT for short). This implies that they are very well applicable to the measurement of change (Moelands et al., 1989; cf. Rogosa & Willett, 1985). With the introduction of the common core curriculum in secondary education a similar system might be introduced based on the tests to assess progress of individual pupils with respect to the common standards. So from here on, the availablity is assumed of a set of reliable and valid observations on the achievement of each pupil i in school j on t different time points ($t > 2$): θ_{tij} with θ being a random variable with cumulative scaling properties. What we have here is the combination of a repeated measueres design (within subjects design) with the nesting of pupils within a given school: the data are ordered hierarchically. In the statistical model to be presented, the basic modelling logic is that one wants to make full use of the data. As each sampling unit is sampled from a population, data from other sampling units (other time points, other pupils, other schools) will be used to improve estimates of pupil achievement, pupil progress, school output and school progress respec-

tively by applying Empirical Bayes techniques. The model described below is an application of a three-level model introduced by Raudenbush (1989, 1990) to study organizational effects on individual growth.

Since achievement is dependent on pupil characteristics like aptitude and sex the following prediction logic is used to model pupil progress (cf. Bryk & Raudenbush, 1987, p.148):

$$(1)\ \underline{\theta}_{tij} = \underline{\pi}_{0ij} + \underline{\pi}_{1ij}t + \underline{\pi}_{2ij}t^2 + \ldots + \underline{\pi}_{(K-1)ij}t^{K-1} + \underline{r}_{tij}$$

in which $\underline{\theta}_{tij}$ is the ability-score of pupil i in school j at timepoint t, $\underline{\pi}_{0ij}$ represents the base achievement level for this pupil at the start and t (= 0,1,2,3,...,T) is time elapsed since the first measurement, with the remainder of the expression describing the growth curve characteristics, and the residual term $\underline{r}_{tij}$. According to the design of the CITO pupil monitoring system there are 14 measurement points in time (two tests taken at half a years' intervals each grade from grade 2 up to grade 8). Expression (1) represents a polynomial of degree K-1, that may be reduced to a smaller degree depending on the significance tests for $\underline{\pi}_{kij} = 0$ with k = 0,1,...,K-1.

In a graphical representation of the polynomial of the first degree we may find the growth curves as depicted in Figure 1:

In this example the two pupils differ in their initial achievement as well as in their growth rate. Explaining these two differences can be achieved in the second equation:

$$(2)\ \underline{\pi}_{kij} = \underline{\pi}_{k0j} + \underline{\beta}_{kj}x_{ij} + \underline{e}_{kij},$$

which says that the variation in the growth curve characteristics $\underline{\pi}_{kij}$, with k = 1,2,..., K-1, of a pupil i in school j can be accounted for by a pupil variable like 'aptitude': smart pupils, for

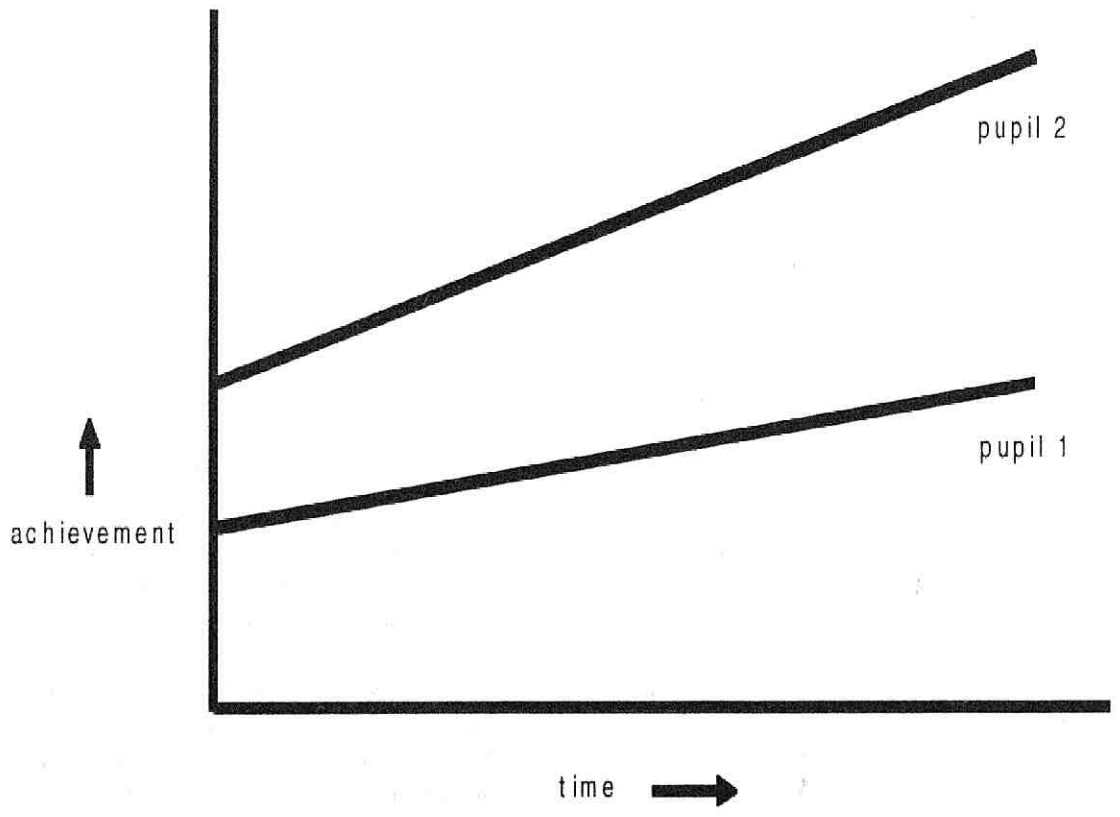


Fig. 1. Differences in growth in ability.

instance, may have a high initial achievement status and a steep growth curve. It is important to stress this point, because this specification is necessary to distinguish school from pupil effects. What is special about this formulation is the allowance of random parameters in this between pupil model: $\underline{\pi}_{k0j}$ are the common elements of the growth curve characteristics of all pupils in school j; furthermore the regression of pupil growth curve characteristics on pupil background variables may vary between schools. Aptitude, for instance, may affect the growth of pupils in some schools (e.g. schools using ATI-principles) to a much lesser degree than in other schools. In order to keep the discussion as simple as possible, it is assumed that the random variable $\underline{\pi}_{k0j}$ is of prime interest. Furthermore it is assumed (for simplicity sake) that the variance of the parameters $\underline{\beta}_{kj}$ is 0 ($\underline{\beta}_{kj} = \underline{\beta}_k$ for k = 0,1,..,K-1). The $\underline{\pi}_{k0j}$ are school specific, and they can be expressed as:

(3) $\underline{\pi}_{k0j} = \underline{\pi}_{k0} + \gamma_k z_j + \underline{u}_{kj}$

In words: the mean growth curve characteristics of school j consist of school specific deviations $\underline{u}_{kj}$ from the growth curve characteristics across all schools $\underline{\pi}_{k0}$. The mean growth curve characteristics of a school can be explained by a school level variable z_j like 'denomination' for instance.

How can this hierarchical model be used for self-evaluation?
First of all, things may be simplified by assuming that there are no background variables. Then one is left with the following simple expressions:

(4) $\underline{\pi}_{k0j} = \underline{\pi}_{k0} + \underline{u}_{kj}$

which indicates how to draw a line through all the observations of all schools, and how individual schools deviate from this line.

(5) $\underline{\pi}_{kij} = \underline{\pi}_{k0j} + \underline{e}_{kij}$,

which indicates how to draw a line through the observations for all pupils in a given school, and how individual pupils deviate from this line.

(6) $\underline{\theta}_{tij} = \underline{\pi}_{0ij} + \underline{\pi}_{1ij} t + \underline{r}_{tij}$

which indicates how to draw a line through the observations of each pupil in a given school, and how each observation in time deviates from this line.

Interesting for self-evaluation purposes are the $\underline{\pi}_{k0j}$: how good does a school perform compared with other schools? That full use is made of the data can be seen in the way the $\underline{\pi}_{k0j}$ are estimated using Empirical Bayes methods (cf. Raudenbush & Bryk, 1986, p. 150):

(7) $\underline{\pi}^*_{\emptyset j} = w_j \underline{\pi}_{k0j} + (1-w_j)\underline{\pi}^*_{\emptyset}$

The w_j's are weights, that are based on the number of observations within schools. The growth-trajectory for school j is estimated as a weighted sum of the OLS-estimate of this trajectory ($\underline{\pi}^*_{k\emptyset j}$) and the Empirical Bayes estimate of the mean growth trajectory across schools. ($\pi^*_{\emptyset}$)For instance, the fewer observations there are for school j the more the Emprical Bayes estimate is based on the average growth trojectory.

SELF-EVALUATION BY STUDYING GROWTH IN ACHIEVEMENT

To illustrate the self-evaluation procedure as outlined above, data were used from a current study on self-evaluation, in which pupil-, class, and schoolperformance are fed beack to the schools to study their effectiveness as compared with a control group that does not receive feedback. Data are available on 373 pupils in 22 groups in lower general secondary schools (mavo). Each pupil was tested three times: at the start of grade 1, at the end of grade 1 and at the end of grade 2. The tests were vertically equated and the ability scores for each pupil were calculated using the one parameter logistic model (Verhelst, 1992). For this example K=2, in other words linear growth is assumed. The data used in this example are the data that were fed back to the schools. The assessment of the effects of the experiment will be based on a test to be administrated to the pupils at the end of grade 3, and will be reported by the end of 1994.
Some descriptive statistics of the variables are presented in Table 2.

There are 8 different ways to estimate the performance of a school.

A.1. The most simple statistic is the mean (raw) test score of all pupils within a school at time t=2 ($\hat{y}_{2.j}$).

A.2. The next one is based on the mean ability score of the pupils at time t=2 ($\hat{\theta}_{2.j}$).

A.3. The third is based on an Empirical Bayes estimate of the score of the pupils after two years of schooling, making full use of prior knowledge on the achievement of the pupils at the two earlier timepoints $((\hat{\pi}_{000} + \hat{u}_{00j}) + 2*(\hat{\pi}_{100} + \hat{u}_{10j}))$, in which $\hat{\pi}_{000} + \hat{u}_{00j}$ is the estimated base achievement level for school j at time t=o, and $\hat{\pi}_{100} + \hat{u}_{10j}$ is the estimated linear rate of change between two consecutive time points.

These first three indicators (A.1. through A.3.) are based on the achievement of pupils after two years of schooling. These indicators can be especially useful when comparing achievement levels with absolute standards, assuming that these are defined on the mathematics ability scale as used in the tests. In this respect indicators A.2. and A.3. are superior to A.1. since they project the scores of the pupils on this scale in a psychometrically correct way. A serious drawback of these indicators A.1. through A.3., however, is that they do not reflect solely the quality of (outputs of) the educational practices, but input differences between schools as well. The next indicator takes some of these input differences into account.

B.1. The fourth estimate of the performance of a school is as A.3. but with a correction for initial differences between pupils with respect to sex and aptitude and with a correction for aptitude effects on growth. This performance index may be conceived as the mean overachievement of the pupils in a school.

Although indicator B.1. does away with the drawbacks of gross effect measures, one might still argue that a more refined and fair measure should take into account input differences between schools with regard to mathematics achievement. This can be achieved by using "progress" indicators.

C.1. The most simple statistic to assess progress is the difference between the mean achievement score at point 2 and point 0 ($\hat{y}_{2.j} - \hat{y}_{0.j}$).

C.2. Somewhat more sophisticated is the same statistic, but now based on the ability scores ($\hat{\theta}_{2.j} - \hat{\theta}_{\emptyset.j}$).

Table 2. Means and standard deviations of the variables.

variable	mean	s.d.
mathematics ability at t=0	-.23	.40
mathematics ability at t=1	.02	.28
mathematics ability at t=2	.21	.38
sex (0= boy; 1= girl)	.52	
aptitude	0	7.81

C.3. The Empirical Bayes estimate of the mean growth in a school is the indicator using information at the three timepoints ($2*(\hat{\pi}_{100} + \hat{u}_{10j})$).

C.4. The last indicator is once again the Empirical Bayes estimate of the mean growth, but now once again with the same correction as in B.1. (for sex and aptitude effects on initial status and also for aptitude effects on growth).

The hierarchical linear model used in A.3. and C.3. is:

(8) $\underline{\theta}_{tij} = \underline{\pi}_{0ij} + \underline{\pi}_{1ij}t + \underline{r}_{tij}$
(9) $\underline{\pi}_{0ij} = \underline{\pi}_{00j} + \underline{e}_{0ij}$
(10) $\underline{\pi}_{1ij} = \underline{\pi}_{10j} + \underline{e}_{1ij}$
(11) $\underline{\pi}_{00j} = \underline{\pi}_{000} + \underline{u}_{00j}$
(12) $\underline{\pi}_{10j} = \underline{\pi}_{100} + \underline{u}_{10j}$

In estimating B.1. and C.4. sex and aptitude are introduced as predictors of the initial ability score. Moreover aptitude is also introduced to predict differences in growth rates between pupils. Expressions (9) and (10) therefore are altered into:

(13) $\underline{\pi}_{0ij} = \underline{\pi}_{00j} + \beta_1 sex_{ij} + \beta_2 IQ_{ij} + \underline{e}_{0ij}$
(14) $\underline{\pi}_{1ij} = \underline{\pi}_{10j} + \beta_3 IQ_{ij} + \underline{e}_{1ij}$

The results of the analyses are presented in Table 3, where, for simplicity's sake, the mean is subtracted from the estimates.

Table 3. Estimates of school effects (ranks between brackets), and Spearman correlations (in the last row) between C.4. and the other estimates.

group	net growth C.4.		gross growth C.3.		ability diff. C.2.		raw diff. C.1.		t=2 net EB B.1.		t=2 gross EB A.3.		t=2 ability A.2.		t=2 raw A.1.	
3021	-.21	(1)	-.21	(1)	-.28	(1)	-5.77	(1)	-.06	(8)	-.07	(8)	-.05	(9)	-1.80	(10)
5042	-.16	(2)	-.15	(2)	-.20	(3)	-5.13	(2)	-.11	(4)	-.15	(4)	-.21	(5)	-4.95	(4)
3022	-.15	(3)	-.14	(3)	-.20	(2)	-4.14	(6)	-.05	(9)	-.11	(7)	-.11	(6)	-3.16	(6)
2013	-.11	(4)	-.11	(4)	-.16	(4)	-2.92	(8)	.01	(12)	.02	(13)	-.02	(12)	-1.99	(8)
2231	-.11	(5)	-.10	(5)	-.12	(6)	-3.32	(7)	-.11	(5)	-.14	(6)	-.09	(7)	-2.48	(7)
5021	-.10	(6)	-.09	(7)	-.11	(7)	-5.09	(3)	-.24	(1)	-.32	(1)	-.40	(1)	-10.07	(1)
5041	-.10	(7)	-.09	(8)	-.11	(8)	-2.65	(9)	-.15	(2)	-.19	(2)	-.23	(3)	-4.32	(5)
5141	-.09	(8)	-.10	(6)	-.13	(5)	-2.32	(11)	-.10	(7)	-.05	(9)	-.07	(8)	-.42	(12)
2371	-.04	(9)	-.06	(9)	-.07	(9)	3.47	(18)	.26	(22)	.36	(22)	.46	(22)	11.35	(22)
5142	-.04	(10)	-.04	(10)	-.05	(10)	-2.64	(10)	-.05	(10)	-.03	(11)	-.05	(10)	-1.86	(9)
5022	-.03	(11)	-.03	(12)	-.04	(12)	-4.17	(5)	-.11	(6)	-.15	(3)	-.23	(4)	-6.51	(3)
2232	-.02	(12)	-.03	(11)	-.04	(11)	2.36	(15)	.01	(13)	.05	(14)	.12	(18)	4.41	(18)
2012	.00	(13)	.01	(13)	.01	(13)	2.26	(14)	.15	(20)	.12	(19)	.17	(19)	4.64	(19)
5081	.01	(14)	.02	(14)	.02	(14)	1.44	(12)	.00	(11)	-.04	(10)	-.05	(11)	.12	(13)
2011	.02	(15)	.02	(15)	.04	(15)	-4.50	(4)	-.12	(3)	-.14	(5)	-.29	(2)	-9.85	(2)
2373	.04	(16)	.03	(16)	.04	(16)	2.21	(13)	.06	(16)	.09	(17)	.10	(17)	3.05	(15)
2401	.08	(17)	.07	(17)	.08	(17)	5.25	(20)	.25	(21)	.30	(21)	.40	(21)	10.05	(21)
2372	.12	(18)	.11	(18)	.13	(18)	4.50	(19)	.13	(19)	.18	(20)	.21	(20)	6.10	(20)
5082	.15	(19)	.15	(19)	.20	(19)	3.19	(16)	.04	(15)	.07	(15)	.07	(14)	.78	(14)
2402	.20	(20)	.20	(20)	.27	(20)	3.32	(17)	.03	(14)	.00	(12)	.01	(13)	-.66	(11)
5111	.26	(21)	.25	(21)	.32	(21)	6.00	(21)	.08	(18)	.09	(16)	.10	(16)	3.21	(16)
5112	.30	(22)	.29	(22)	.38	(22)	8.57	(22)	.07	(17)	.12	(18)	.09	(15)	4.33	(17)
Corr.	1.00		1.00		.99		.81		.60		.59		.53		.54	

The 22 groups are ordered on the basis of their scores on indicator C.4., being an estimate of the growth in grade 1 and 2, corrected for sex and aptitude (defense mechanisms like "we don't have very bright pupils" are therefore ruled out). C.4. and C.3. almost lead to the same ranking of the 22 groups.

The ranking is also very similar (.99) with the more convenient indicator C.2., but that is only so because there are only three timepoints in the example! When more timepoints are added the difference in ranking between C.2. on the one hand and C.3. and C.4. will become more manifest. Notice also that C.4. and C.3. have less variation than C.2., which is the direct consequence of the Empirical Bayes approach, that works like "regression to the mean": the less information there is for a certain school the more its estimate is shrunk toward the grand average.

Greater differences in ranking occur when using the (easy to compute) difference score of the two raw test scores (the correlation now reduces to .81).

Notice, for instance, the misleading information that is presented about group 2011 (at rank 15 at C.4.), that is performing on average according to C.4., but is doing worse on C.1 (at rank 4). The opposite occurs for group 2371; that is at rank 9 at C.4. and at rank 18 at C.1. A complete other ranking occurs when applying B.1. and A.1., A.2., or A.3., as can be readily deduced from the magnitudes of the correlation coefficients (.60, .59, .53, and .54 respectively).

Which of these 8 indicators of school performance should be preferred in case of school self-evaluation? It should be the one that conceptually reflects the effort that the school has made to make its pupils achieve. Stated otherwise, the one indicator that is of primary interest in case of self-evaluation is similar to the preferred criterium in school effectiveness studies, and to the one that one would use in informing parents in a fair way about the school of their children. Our first contention is that one would refrain from using cross-sectional data only, since these may reflect input differences between schools. For school self-evaluation purposes it is of course interesting to find out whether a school meets a certain standard, e.g. a cut-off score on the scale. For this reason A.2. is a better indicator than A.1., the one that is most likely to be used. Chance fluctuations are always possible, and for this reason indicator A.3. is a statistical improvement. The indicators that conceptually come closest to a fair definition of a school effect are the indicators that reflect the progress that pupils make, with, mutatis mutandis, as limitations and advantages the ones just mentioned for the indicators A.1. through A.3. The real problem with these progress indicators is the assessment of the initial achievement level (the base), since in some cases it may not make much sense to test pupils, for instance, for their initial knowledge of Greek language. For this reason one may assess the base level somewhat later than at school entrance, or one may choose to use the overachievement indicator B.1. instead. This is the one proposed by Willms (1992).

Our second contention is that there are a lot of psychometric requirements to be imposed on achievement tests (see also: Seltzer, Frank & Bryk, 1994). Advanced pupil monitoring systems meet these standards. These, together with the conceptual deliberations regarding a fair definition of a school effect, lead to the conclusions that pupil monitoring systems provide a sound basis for school monitoring systems. This brings us to the question of how schools can find those grades and / or teachers within a school that mostly affect the effectiveness of a school in a certain subject area.

Statistically this can be achieved by introducing an intermediate level in the hierarchy between pupils and schools, assuming that the tests are constructed in such a way that, on average, linear growth within and across grades is the case. (This latter assumption can be relaxed by introducing a fixed grade-effect on growth). Comparing each grade in a school with all comparable grades in a sample actually may lead to the identification of poorly operating classes and educational practices. In this way schools learn where to locate the leverage point for change. If, moreover, schools also use an indicator system for assessing these educational practices (e.g. instruments to measure "allocated learning time", or grouping practices), links can be made between the quality indicators of process and output. The procedure to follow in this latter case depends on the characteristics of the self-evaluation system as implemented. First of all, if schools perform below average on the achievement tests as well as on some of the process

indicators, this information can be used to detect those practices where improvement is possible. A more sophisticated approach is, that schools, knowing the production function (in terms of a regression model), decide on the basis of their observed achievement gain, process scores, and the (known) regression coefficients, how improvement can be optimized by implementing change in some of the process characteristics. Lastly, it is possible, that schools seek for internal points of reference (comparing one class with other classes, or making comparisons in time), thus trying to find leverage points for change.

USE AND GENERALIZATION

Process-product relations

When a school would have a full-fledged pupil monitoring system in place, the condition of having primary output data for self-evaluation purposes would be met in a most convincing way.

Measurement of the key-criteria variables, however, is only one vital aspect of evaluation, the other being the internally and externally valid causal attribution of outcomes to antecedent conditions. In other words the natural interest in school self-evaluation will be directed at process-product relationships. Questions like: why do the results in a certain year fall back with respect to the previous year: why do the results with method A appear to be better than with method B, etc.

It is beyond the scope of this article to discuss this issue in any detail, but two lines of approach shall be mentioned that are possible when one deals with the type of output information that comes available from pupil monitoring systems.

First, within a school, teaching conditions could be compared. Particularly when age groups would be in parallel classes (e.g. two first grades, two second grades, etc.), they would form a natural basis to conduct "experiments of nature" (exploiting natural variance) or even planned quasi-experiments. In the latter case Campbell's famous dictum of "reforms as experiments" could be generalized to the within schools level.

In the absence of parallel classes within subject designs as described could be employed to check whether certain innovations or "interruptions" would show up in the achievement patterns.

Secondly, organizational measures at the school level, or school conditions in general, could be similarly assessed by means of within subject designs.

Thirdly, when several schools, for instance within a certain neighbourhood, would employ the same pupil monitoring system and coordinate data collection (i.e. measure at the same levels at the same points in time), all types of causal analysis, with respect to both variables at school and class level, that we know from empirical school effectiveness studies, would be applicable.

In this way networks of schools might start to function as committees of organizations that learn from each other.

Table 4. Prerequisites for output oriented school self-evaluation.

- availability of tests, preferably meeting item-response models
- user friendly approaches to test administration
- facilities (preferably computerized) for data storage, analysis and retrieval
- manpower for internal control and administrative tasks surrounding a monitoring system
- consensus on the use of a monitoring system among staff
- support of school management
- coordinated use of a monitoring system
- proficiency in data analysis, quasi-experimental designs
- meetings of staff and management to discuss reports and to discuss courses of action

Practical prerequisites

School self-evaluation as described in this article has some utopian features. We realize that in actual practice a lot of barriers need to be taken. So, in order to avoid the accusation of being unrealistic, it is perhaps good to summarize the prerequisites for the type of school self-evaluation that has been the center of attention in this article. This is shown in Table 4.

It is quite evident that several of these functions would require the support by external agencies, such as institutes for test development or educational research specialists.

Self-evaluation using product data should take into account the psychometric, statistical, and conceptual deliberations concerning the measurement of school effects. Further developments in computerized testing and in computer software for use in pupil monitoring systems might also lead to the growth of a generation of software for school monitoring systems that minimizes external support.

REFERENCES

Bluhm, H.P. & Visscher, A.J. (1990). Administrative computing in the USA and the Netherlands: implications for other countries. *School Organisation, 10* (1), 107-117.

Bryk, A.S. & Raudenbush S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147-158.

Gillijns, P. & Verhoeven, L. (1991). Naar een leerlingvolgsysteem voor het basisonderwijs. *Pedagogische Studiën, 68*, 216-230.

Hopkins, D. (1987). *Doing School Based Review*. Leuven, Belgium: ACCO.

Leeuw, A.C.J. de (1988). *Organisaties: management, analyse, ontwerp en verandering. Een systeemvisie*. Assen: Van Gorcum.

Moelands, A.H.J., Mommers, M.J.C. & Oud, J.H.L. (1989). Leerlingvolgsystemen verklaard en vergeleken. *Schoolbegeleiding, 7* (26), 19-28.

Oakes, J. (1987). *Conceptual and measurement issues in the construction of school quality indicators*. Paper presented at the AERA Annual Meeting. Washington D.C.

Raudenbush, S.W. (1989). The analysis of longitudinal, multilevel data. *International Journal of Educational Research, 13*, 721-740.

Raudenbush, S.W. (1990). Dynamic models of school and teacher effects on student learning. In: P. van den Eeden, J. Hox & J. Hauer (eds.). *Theory and model in multilevel research: convergence or divergence?* (pp. 99-125). Amsterdam: SISWO.

Raudenbush, S.W. & Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*, 1-17.

Rogosa, D.R. & Willett, J.B. (1985). Understanding correlates of change by modelling individual differences in growth. *Psychometrika, 90*, 726-748.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. In D. Reynolds & B.P.M. Creemers (eds.), *School Effectiveness and School Improvement* (pp. 61-80). Lisse: Swets & Zeitlinger.

Seltzer, M.H., Frank, K.A. & Bryk, A.S. (1994). The metric matters: The sensitivity of conclusions about growth in student achievement to choice of metric. *Educational Evaluation and Policy Analysis, 16* (1), 41-49

Teauber, R.C. (ed.) (1987). Education data system redesign. *International Journal of Educational Research, 11* (4), 391-511.

Verhelst, N.D. (1992). *Het eenparameter logistch model (OPLM): een theoretische inleiding en een handleiding bij het computerprogramma*. (OPD-memorandum nr. 3). Arnhem: CITO.

Voogt, J.C. (1989). *Scholen doorgelicht. Een studie over schooldiagnose*. De Lier: Academisch Bocken Centrum.

Willms, J.D. (1992). *Monitoring school performance: a guide for educators*. Washington, D.C.: Falmer.

Manuscript ontvangen 28-4-1944
Definitieve versie ontvangen 2-9-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 2, pp. 165-177

Studiesucces van buitenlandse studenten

Studievoortgang van buitenlandse studenten in het technisch-wetenschappelijk onderwijs geanalyseerd

H.J. Smid, A. Verweij, W. Jochems en J. Snippe[1]
Technische Universiteit Delft, sectie Didactiek en Onderwijsontwikkeling

ABSTRACT

Institutions of higher education are faced with an increasing number of foreign students, which may raise the issue of the latters' academic performance. In a study at Delft University of Technology, study progress of foreign students has been compared with study progress of all students. Furthermore, data on foreign students' achievement have been analyzed in order to determine what factors contribute to foreign students' progress. From data collected at three faculties, the propaedeutic output of foreign students appears to be the same as for other students, although foreign students need more time to pass the propaedeutic examination. Selection in the countries of origin appears to be a successful way of excluding students who are not suited to study abroad. In addition, foreign students for whom successful completion of their studies will be problematic appear to be detected at an early stage. This makes it possible to provide counselling or to advise a timely change of studies.

1. INLEIDING

Ten gevolge van internationalisering is er sprake van een toenemende instroom van buitenlandse studenten. Niet alleen stijgt het aantal buitenlanders dat aan korte postdoctorale cursussen deelneemt, maar ook het aantal dat in het reguliere Nederlandse hoger onderwijs instroomt is groeiende. Volgens de laatst beschikbare gegevens hebben ongeveer 5500 studenten in het WO een buitenlandse nationaliteit (zie Van der Wende, 1992); dat is circa 3% van de studenten in het WO. Bij technische wetenschappen ligt dat percentage hoger; bij de TU Delft bijvoorbeeld is ruim 10% van de eerstejaars-instroom van buitenlandse afkomst, hetgeen bijna 300 buitenlanders per jaar betekent die tot een studie worden toegelaten (Statistisch Jaarboek, 1993). Het is duidelijk dat het hier om een substantiële groep studenten gaat. Het overheidsbeleid is bovendien gericht op een verdere toename van het aantal buitenlandse studenten in het hoger onderwijs.

Studenten van buitenlandse afkomst die hier komen studeren, ondervinden nogal wat problemen. Naast diverse praktische problemen zoals huisvesting en het verkrijgen van een verblijfsvergunning (Van der Wende, 1992; Smit, 1994; Snippe, 1994), vormt de Nederlandse taal een groot obstakel. Aangezien vrijwel al het onderwijs in het Nederlands plaatsvindt, zal de buitenlandse aspirant-student moeten kunnen aantonen dat hij deze taal in redelijke mate beheerst. Hij legt hiertoe een Nederlandse taalvaardigheidstoets af. Als voorbereiding op een studie in het Nederlandse hoger onderwijs kan bij de meeste instellingen voor hoger onderwijs een cursus Nederlands worden gevolgd.

Een ander probleem heeft te maken met de vereiste voorkennis. Om voor een volledige studie aan een instelling voor hoger onderwijs te worden toegelaten dient een buitenlandse aspirant-student een diploma van het voortgezet onderwijs te kunnen overleggen dat in eigen land toegang verschaft tot het hoger onderwijs. Op basis van de waardering van dit diploma door de betreffende universiteit of door de Nuffic wordt het diploma als geldig bewijs van toelating beschouwd. Voor studenten afkomstig uit een land dat tot het Europees verdrag hoort, wordt het

[1] Adres: TU Delft, sectie Didactiek en Onderwijsontwikkeling, Kanaalweg 2b, 2628 EB Delft.

middelbare schooldiploma zondermeer erkend als middel tot toelating tot het Nederlandse hoger onderwijs. Daarnaast stellen diverse instellingen voor hoger onderwijs aanvullende eisen aan de voorkennis van de aspirant-studenten. Zo moeten studenten die aan de TU Delft willen gaan studeren, kunnen aantonen dat ze over de vereiste kennis van wiskunde en natuurkunde beschikken, dat wil zeggen op een niveau dat equivalent is aan dat van VWO-instromers. Personen die niet over voldoende kennis op deze terreinen beschikken, dienen aanvullende cursussen te volgen. Veel universiteiten bieden buitenlandse studenten faciliteiten om deficiënties in de vooropleiding weg te werken. Een student, afkomstig uit een land dat bij het Europees verdrag is aangesloten, wordt toegelaten op grond van een diploma voortgezet onderwijs zonder dat wordt nagegaan of diens voorkennis toereikend is.

In verband met het toenemend aantal buitenlandse studenten en tegelijkertijd de toenemende druk op universiteiten om het rendement van hun studies te verhogen is empirisch gefundeerde informatie over het studiesucces van buitenlandse studenten van groot belang. In dit artikel wordt verslag gedaan van een onderzoek waarin de studievoortgang van buitenlandse studenten is vergeleken met die van Nederlandse studenten en waarin is nagegaan welke factoren het studiesucces van buitenlandse studenten beïnvloeden. Op basis van deze empirische gegevens worden enkele beleidsmaatregelen ter bevordering van het studiesucces van buitenlandse studenten besproken en de mogelijke kans op succes van dergelijke maatregelen.

2. THEORETISCHE ACHTERGROND

Hoewel het aantal buitenlandse studenten dat hier een volledige studie in het hoger onderwijs komt volgen, is toegenomen, is er weinig empirisch onderzoek verricht naar studiesucces van buitenlandse studenten.

Admiraal & Boogaard (1994) hebben de studieresultaten van Nederlandse en buitenlandse studenten bij drie Nederlandse universiteiten met elkaar vergeleken. Het gaat hier om de Landbouwuniversiteit Wageningen, de Technische Universiteit Eindhoven en de Rijksuniversiteit Utrecht. Buitenlandse studenten zijn in dit onderzoek gedefinieerd als studenten die in het buitenland zijn geboren én een buitenlandse nationaliteit hebben én tevens in het bezit zijn van een buitenlands diploma van voortgezet onderwijs. Uit het onderzoek blijkt dat buitenlandse studenten iets langer over de propaedeusefase doen en vervolgens iets korter over de doctoraalfase in vergelijking met hun Nederlandse mede-studenten. Over het geheel genomen is de verhouding van de studieduur die nodig is om het propaedeusediploma te halen en de totale studieduur voor buitenlandse studenten anders dan voor Nederlandse studenten. Zo gebruiken buitenlandse studenten ongeveer eenderde deel van hun studietijd voor het behalen van hun propaedeusediploma, terwijl dit voor Nederlandse studenten ongeveer een kwart is van hun totale studieduur.

De vergelijking is echter problematisch waar buitenlandse studenten een studierichting kiezen die verwantschap heeft met het land van herkomst, bijvoorbeeld Turken die Turks gaan studeren. In die studierichting behalen ze over het algemeen betere studieresultaten dan de Nederlandse studenten. Dit maakt echter dat een vergelijking tussen Nederlandse en buitenlandse studenten over studierichtingen heen niet geheel zonder problemen is. Uit het onderzoek wordt verder niet duidelijk door welke factoren de verschillen in studiesucces kunnen worden verklaard, anders dan studierichting.

Ook in de internationale literatuur zijn empirische gegevens over studievoortgang van buitenlandse studenten in het hoger onderwijs schaars. Op grond van de literatuur die beschikbaar is, kunnen de volgende veronderstellingen worden geopperd.

De belangrijkste predictoren voor studiesucces van studenten in het hoger onderwijs zijn enerzijds bekwaamheid en inzet van de student en anderzijds de kwaliteit van het onderwijs (Fraser et al. 1987; Walberg, 1986). Wanneer bekwaamheid wordt geoperationaliseerd als voorkennis in de zin van eindexamencijfers van de vooropleiding, dan is een correlatie met studiesucces te verwachten van rond de .50 (Wellens, 1980). De correlatie kan nog iets hoger worden

wanneer als indicator voor voorkennis de voorafgaande studieprestaties in dezelfde studie worden genomen (Van Overwalle, 1985). Inzet, gemeten als de tijd die de student daadwerkelijk aan de studie besteedt, heeft een correlatie van circa .40 met studiesucces. Op deze globale schets van het verklaringsmodel van studiesucces is nog wel het een en ander aan te merken. Zo wordt de sterkte van verbanden beïnvloed door de wijze waarop de variabelen worden gemeten. Voor de operationalisatie van studiesucces wordt vaak het gemiddelde tentamencijfer in de eerste periode van de studie gekozen of, voor Amerikaanse studies, de GPA (Grade Point Average; Graham, 1987). Het nadeel hiervan is dat het tijdsinterval waarover een gemiddelde wordt berekend, sterk kan variëren, evenals het aantal tentamens waarop dat gemiddelde is gebaseerd. Studiesucces kan ook worden gedefinieerd als het totaal aantal punten dat men in een studiejaar heeft behaald of als het tempo waarmee men de studie doorloopt. Bij deze laatste operationalisatie moet men zich ervan vergewissen dat het studieprogramma geen belemmeringen biedt voor studenten om in eigen tempo het programma te doorlopen.

Voor het opstellen van een verklaringsmodel voor studiesucces van buitenlandse studenten dient het model te worden uitgebreid met de factor beheersing van de vreemde taal waarin het onderwijs plaatsvindt (Jochems, 1991). Voor zover er empirische studies beschikbaar zijn op het gebied van studeren in een vreemde taal, betreffen ze buitenlandse studenten in Engelstalig onderwijs bij sterk uiteenlopende opleidingen. In dergelijke studies wordt nagegaan of er een significant verband bestaat tussen de vreemde-taalbeheersing -zoals gemeten met een taaltest- en studiesucces. Het studiesucces wordt veelal geoperationaliseerd als Grade Point Average. Uit de resultaten blijkt dat in technische en natuurwetenschappelijke studies de vreemde-taalbeheersing een lagere correlatie met studiesucces vertoont dan bij sociaal-wetenschappelijke studies (Sharon, 1972; Light et al., 1987). De verklaring hiervoor zou kunnen zijn dat taal in technische studies een minder belangrijke rol speelt. Verder lijkt de factor taalbeheersing minder belangrijk ten aanzien van het realiseren van studievoortgang te worden naarmate buitenlandse studenten de vreemde taal in kwestie beter beheersen. Zo rapporteren Ho en Spinks (1985) na één jaar studie een correlatie van r=.30, welke na drie jaar studie was afgenomen tot r=.22. Wat de onderlinge bijdrage van de factoren bekwaamheid, inzet en vreemde-taalbeheersing betreft, ziet het ernaar uit dat in technische en natuurwetenschappelijke opleidingen beheersing van wiskunde een betere voorspeller van studiesucces is dan vreemde-taalbeheersing (Ayers & Peters, 1977; Sharon, 1972). Voor andere opleidingen echter is beheersing van de vreemde taal in kwestie een betere voorspeller van studiesucces.

In de hier aangehaalde studies is taalvaardigheid gemeten met de TOEFL (Test Of English as a Foreign Language). Het valt echter te betwijfelen of deze test alle aspecten van taalvaardigheden in kaart brengt die van belang zijn voor studiesucces en dus een goede voorspeller vormt voor studiesucces (Light et al., 1987).

In het hier beschreven onderzoek hebben de volgende onderzoeksvragen centraal gestaan:

1. Hoe is het studiesucces van buitenlandse studenten in het reguliere Nederlandse hoger onderwijs in vergelijking met Nederlandse collega-studenten?
2. Welke factoren spelen een rol bij de verklaring van verschillen in studiesucces van buitenlandse studenten?

Deze vragen zijn onderzocht bij een drietal faculteiten binnen het technisch hoger onderwijs. Het postdoctorale onderwijs is buiten beschouwing gebleven vanwege het specifieke karakter van dergelijk onderwijs.

3. ONDERZOEKSOPZET

Als onderdeel van een onderzoek naar het studiesucces van buitenlandse studenten aan de Technische Universiteit te Delft zijn studieresultaten van buitenlandse studenten aan de faculteiten Elektrotechniek, Werktuigbouwkunde en Informatica van deze universiteit onderzocht (Jochems et al., 1993; Jochems et al., 1994). Buitenlandse studenten zijn gedefinieerd als stu-

denten die niet in het bezit zijn van een Nederlands diploma dat toegang verschaft tot het hoger onderwijs en die het Nederlands bij aanvraag tot toelating niet beheersen. Volgens deze definitie worden Vlamingen dus niet tot de buitenlandse studenten gerekend en Surinamers ook niet.

In het onderzoek zijn alleen resultaten over de propaedeusefase verzameld. De gedachte hierachter is dat de relatie tussen bepaalde achtergrondvariabelen en studiesucces van buitenlandse studenten het meest pregnant naar voren zal komen in de beginfase van de studie. Bovendien zijn doktoraalprogramma's zowel binnen als tussen de faculteiten moeilijker onderling te vergelijken. In het onderzoek is de factor studiesucces geoperationaliseerd in drie variabelen: a) het percentage studenten dat het propaedeusediploma behaalt, ook wel het propaedeuserendement genoemd; b) de studieduur tot aan het behalen van het P-diploma; c) het aantal behaalde studiepunten op diverse momenten in de P-fase. Als achtergrondvariabelen zijn in het onderzoek opgenomen geboorteland, geslacht, leeftijd en taalvaardigheid. De variabele 'taalvaardigheid' is geoperationaliseerd als de score op de toelatingstoets voor Nederlandse taalvaardigheid waarin ook technisch Nederlands is verdisconteerd. De variabele 'kennis van de exacte vakken' kon in dit onderzoek helaas niet worden opgenomen, hoewel het een relevante factor is voor het studiesucces in technische studies. Het probleem is dat niet van alle buitenlandse studenten bekend is wat hun niveau in de exacte vakken zoals wiskunde en natuurkunde is. Studenten die afkomstig zijn uit een land dat onder het Europees verdrag valt, zijn rechtstreeks toelaatbaar en hoeven geen toelatingstoets voor wiskunde en natuurkunde af te leggen. Dit gebeurt wel bij studenten die uit een land komen dat niet onder het Europees Verdrag valt. Deze studenten moeten zowel de wiskundetoets als de natuurkundetoets met goed gevolg hebben afgelegd om tot de studie toegelaten te kunnen worden. Van de onderzochte groep buitenlandse studenten bleek echter geen enkel natuurkundecijfer bewaard te zijn gebleven. In veel gevallen was ook van de wiskundetoets alleen genoteerd wanneer deze voldoende was gemaakt, maar bleek het behaalde cijfer niet meer te achterhalen. Bovendien zijn inhoud en niveau van de wiskundetoets in de loop der tijd niet constant gebleven. Al deze omstandigheden hebben ertoe geleid de variabele 'kennis van exacte vakken' niet in het onderzoek op te nemen.

4. SAMENSTELLING VAN DE STEEKPROEF

De groep buitenlandse studenten heeft de volgende samenstelling. Van alle buitenlandse studenten die in de studiejaren 83/84, 84/85 en 85/86 bij het talenpracticum aan de TU Delft stonden ingeschreven om Nederlands te leren, is nagegaan wie zich vervolgens hebben ingeschreven voor de studierichtingen Elektrotechniek, Informatica en Werktuigbouwkunde in de periode 1983 tot en met 1987. Studenten die één of meer vrijstellingen kregen voor propaedeuse-onderdelen zijn buiten beschouwing gelaten. De steekproef bestaat uit 83 studenten Elektrotechniek, 52 studenten Informatica en 35 studenten Werktuigbouwkunde.

Van deze groep studenten zijn achtergrondgegevens verzameld bij het talenpracticum en studievoortgangsgegevens bij de betreffende faculteiten. De steekproef kan worden beschreven aan de hand van de beschikbare achtergrondgegevens, namelijk geboorteland, geslacht, leeftijd bij aanvang van de studie en beginniveau zoals dat tot uitdrukking komt in de score op de toets taalvaardigheid. Waar dit zinvol is wordt een vergelijking gemaakt met het totale cohort studenten dat zich in die periode heeft ingeschreven bij de drie studierichtingen (n=4755). In dit cohort zijn ook de buitenlandse studenten opgenomen. Voor de vergelijking is namelijk gebruik gemaakt van gegevens zoals die in de centrale studentenadministratie zijn opgenomen en het is hierbij helaas niet mogelijk gebleken de buitenlandse studenten uit het totale cohort weg te laten.

Opvallend is het grote aantal studenten dat uit Indonesië afkomstig is: ze maken iets meer dan de helft uit van alle buitenlandse studenten bij de drie studierichtingen. Dit pleit ervoor om bij analyses onderscheid te maken tussen studenten die wel of niet uit Indonesië afkomstig zijn. Van de groep buitenlandse studenten uit de steekproef is 18% vrouw (n=31). Voor de totale groep studenten die in die periode bij de drie studierichtingen stond ingeschreven (n=4755),

Tabel 1. Verdeling van buitenlandse studenten over de drie studierichtingen naar geboorteland.

studierichting:	n	Indonesië	Noorwegen	Marokko	Iran	overige
Elektrotechniek	83	45%	13%	5%	14%	23%
Informatica	52	46%	13%	23%	2%	15%
Werktuigbouwkunde	35	71%	6%	3%	3%	17%
totaal	170	51%	12%	10%	8%	19%

bedroeg het percentage vrouwen 4%. Het percentage vrouwen onder de buitenlandse studenten is significant hoger dan onder alle eerstejaarsstudenten (inclusief de buitenlanders) die in dezelfde periode aan de studie begonnen ($z=9.22$; $p<.0001$). De gemiddelde leeftijd van de buitenlandse studenten is bij aanvang van de studie 21.0 jaar (s.d. = 3.1). Ter vergelijking: de gemiddelde leeftijd van de Nederlandse TU-student bij aanvang van de studie is 19 jaar.

Om tot een studie toegelaten te kunnen worden dienen buitenlandse studenten een toets taalvaardigheid met goed gevolg te hebben afgelegd. De range van de scores op de taaltoets is 6.0 tot 10.0 (op één decimaal nauwkeurig). In tabel 2 zijn de scores op de taaltoets opgenomen voor de buitenlandse studenten in de drie studierichtingen. Er is hierbij tevens een uitsplitsing gemaakt naar studenten die wel of niet uit Indonesië afkomstig zijn. Van enkele studenten bleken niet alle scores bewaard te zijn gebleven.

De studenten van de drie studierichtingen verschillen niet van elkaar op hun gemiddelde aanvangsniveau qua taalvaardigheid zoals gemeten bij de taaltoets. Er is ook geen verschil tussen studenten die uit Indonesië afkomstig zijn en de niet-Indonesische studenten.

Samenvattend blijkt dat in de steekproef van buitenlandse studenten de Indonesische studenten een speciale plaats innemen. Niet alleen vanwege hun grote vertegenwoordiging in de groep buitenlandse studenten, maar ook vanwege het feit dat de Indonesische studenten bij Elektrotechniek en Werktuigbouwkunde een selectieproces in het eigen land hebben doorlopen. Voor de vergelijking van buitenlandse studenten onderling is het relevant vast te stellen dat er geen significante aanvangsverschillen op het gebied van taalvaardigheid tussen de buitenlandse studenten van de drie faculteiten bestaan. De proportie vrouwen in de groep buitenlandse studenten is groter dan in het totale cohort en de buitenlandse student is gemiddeld 2 jaar ouder dan de gemiddelde eerstejaarsstudent. Het ligt echter niet voor de hand te veronderstellen dat dergelijke verschillen zullen resulteren in verschillen in studiesucces tussen buitenlandse studenten en het totale cohort.

Tabel 2. Scores op de toelatingstoets voor taalvaardigheid van de buitenlandse studenten, uitgesplitst naar studierichting en naar geboorteland.

		taaltoets		
		M	s	n
Studierichting:	Elektrotechniek	7.8	.8	82
	Informatica	7.9	.7	50
	Werktuigbouwkunde	7.9	.7	33
Geboorteland:	Indonesië	7.8	.6	84
	niet-Indonesië	8.0	.8	81
Totaal:		7.9	.7	165

5. RESULTATEN

Studiesucces van buitenlandse studenten in vergelijking met Nederlandse studenten
Allereerst wordt nagegaan in hoeverre de buitenlandse studenten bij de studierichtingen Elektrotechniek, Informatica en Werktuigbouwkunde in hun studiesucces verschillen van de totale groep studenten die in dezelfde periode stond ingeschreven aan deze drie studierichtingen. In tabel 3 is het percentage buitenlandse studenten dat het propaedeusediploma heeft behaald vergeleken met het percentage studenten met een propaedeusediploma van het totale cohort.

Uit de tabel blijkt dat het slaagpercentage voor het P-examen van de buitenlandse studenten iets lager is dan dat van het totale cohort, doch dit verschil is niet significant. Verder blijkt dat het slaagpercentage van de buitenlandse studenten bij Informatica lager ligt dan voor de totale groep Informatica-studenten; dit verschil is significant ($z=-2.72$; $p<.01$). Opvallend is het hoge slaagpercentage van de buitenlandse Werktuigbouwkundestudenten. Hoewel dit slaagpercentage hoger is dan dat van de totale groep Werktuigbouwkundestudenten, is het verschil niet significant.

Vervolgens is nagegaan hoeveel tijd de buitenlandse studenten nodig hebben gehad om hun P-diploma te behalen, in vergelijking met het totale cohort studenten. De resultaten zijn in tabel 4 weergegeven, waarbij uiteraard alleen de studenten zijn opgenomen die hun P-diploma hebben behaald. Uit de tabel blijkt dat buitenlandse studenten meer tijd nodig hebben om hun P-examen te halen in vergelijking met het totale cohort studenten. Opvallend is dat in het totale cohort slechts 2 à 3 % van de studenten langer dan 3 jaar over de propaedeuse doet, terwijl maar

Tabel 3. Propaedeuserendement van de buitenlanders vergeleken met het totale cohort.

	buitenlanders		totale cohort	
studierichting:	n	perc. P	n	perc. P
Elektrotechniek	83	60%	1712	60%
Informatica	52	40%	1279	59%
Werktuigbouwkunde	35	71%	1764	58%
totaal	170	56%	4755	59%

Tabel 4. Studieduur tot het P-examen van buitenlandse studenten, vergeleken met het totale cohort.

	n	P na 1 jaar	P na 2 jaar	P na 3 jaar
Elektrotechniek:				
* buitenlanders	50	20%	70%	94%
* totale cohort	1026	36%	92%	97%
Informatica:				
* buitenlanders	21	0	29%	76%
* totale cohort	749	17%	89%	98%
Werktuigbouwkunde:				
* buitenlanders	25	12%	64%	88%
* totale cohort	1027	24%	89%	97%
Totaal:				
* buitenlanders	96	14%	59%	89%
* totale cohort	2802	27%	90%	97%

liefst bijna een kwart van de buitenlandse Informatica-studenten langer dan drie jaar over het behalen van het P-diploma heeft gedaan. Dit percentage ligt bij Werktuigbouwkunde op 12% en bij Elektrotechniek op 6%. Het percentage studenten dat na 2 jaar het P-examen heeft behaald is bij alle drie studierichtingen in het totale cohort groter dan in de groep buitenlandse studenten. De verschillen tussen de buitenlandse studenten en het totale cohort voor wat betreft het percentage studenten dat het P-diploma binnen twee jaar heeft behaald, zijn zowel binnen de drie studierichtingen als in de totale groep significant ($p<.001$).

Het is niet mogelijk om buitenlandse studenten te vergelijken met het totale cohort op het aantal behaalde studiepunten op diverse momenten tijdens de studie. Er zijn van de studenten die in de jaren '83 tot '87 aan de studie begonnen namelijk geen gegevens over studiepunten bekend.

Relevante factoren voor studiesucces van buitenlandse studenten

Nu het studiesucces van buitenlandse studenten is vergeleken met dat van het totale cohort studenten, is het van belang vast te stellen welke factoren van invloed zijn op het studiesucces van buitenlandse studenten. Hiertoe is nagegaan of buitenlandse studenten onderling in hun studiesucces van elkaar verschillen op de onderscheiden achtergrondkenmerken, namelijk land van herkomst, sexe, leeftijd en taalvaardigheid.

In de eerste plaats is nagegaan of het land van herkomst leidt tot verschillen in studiesucces. Het aantal buitenlandse studenten afkomstig uit een bepaald land is echter te klein om per studierichting verschillen in studiesucces op basis van geboorteland te vergelijken. Aangezien bijna de helft van de buitenlandse studenten afkomstig is uit Indonesië, is wel nagegaan of de Indonesische studenten qua studiesucces verschillen van de overige buitenlandse studenten. Deze gegevens zijn in tabel 5 opgenomen.

Uit tabel 5 is af te lezen dat de slaagpercentages van de Indonesische studenten bij alle studierichtingen afzonderlijk en over het geheel genomen hoger liggen dan bij de niet-Indonesische studenten. De verschillen zijn niet significant. Bij beide groepen is het slaagpercentage bij Informatica het laagst en bij Werktuigbouwkunde het hoogst.

Als tweede achtergrondkenmerk is de factor sexe opgenomen. Het blijkt dat sexe alleen tot verschillen leidt voor wat betreft de studieduur tot het P-examen: de studieduur is voor de mannen gemiddeld 25 maanden (s.d. = 10; n=83) en voor de vrouwen 31 maanden (s.d. = 12; n=13). Dit verschil is alleen bij de faculteit Elektrotechniek significant ($p<.05$), waar de mannen gemiddeld 22 maanden (s.d. = 7; n=43) over hun P-examen doen en de vrouwen 29 maanden (s.d. = 15; n=7). Het gemiddelde slaagpercentage van de mannelijke buitenlandse studenten is 60% (n=139) en bij de vrouwen 42% (n=31). Dit verschil is niet significant. Ook per studierich-

Tabel 5. Studiesucces van Indonesische studenten in vergelijking met niet-Indonesische studenten, uitgedrukt in P-rendement en studieduur in maanden.

studierichting:		n	percentage P	gemidd. studieduur tot P in maanden (s.d.)
Elektrotechniek:	Indonesië	37	68 %	23 (5)
	niet-Indonesië	46	54 %	22 (11)
Informatica:	Indonesië	24	42 %	32 (10)
	niet-Indonesië	28	39 %	30 (13)
Werktuigbouwkunde:	Indonesië	25	76 %	25 (10)
	niet-Indonesië	10	60 %	29 (9)
totaal	Indonesië	86	63 %	26 (9)
	niet-Indonesië	84	50 %	25 (12)

ting ligt het slaagpercentage van de mannen hoger dan van de vrouwen, maar deze verschillen zijn geen van alle significant.

Het derde achtergrondkenmerk -leeftijd- leidt eveneens tot enige verschillen. De studenten die hun P-examen niet hebben gehaald (n=74), waren bij aanvang van de studie gemiddeld 21.6 jaar oud (s.d.=4.1) [zij waren gemiddeld 20.6 jaar oud (s.d.=2.1)] en dat is een jaar ouder dan de studenten die hun P-examen wel hebben gehaald (n=96). Het verschil in aanvangsleeftijd is significant voor de totale groep op .05-niveau en voor de studierichting Elektrotechniek op .01-niveau. Bij Elektrotechniek is het verschil ruim anderhalf jaar.

Vervolgens is nagegaan hoe de relatie tussen taalvaardigheid en studiesucces is. In tabel 8 zijn de scores op de taaltoets opgenomen voor de buitenlandse studenten die wel en zij die niet het P-diploma hebben behaald.

De gemiddelde score op de taaltoets van de personen die het P-examen hebben behaald is hoger dan de personen die het P-examen niet hebben behaald. Dit geldt voor de totale groep, maar ook voor de studierichtingen afzonderlijk. Behalve bij Werktuigbouwkunde, zijn de verschillen in de score op de taaltoets voor de studierichtingen significant ($p<.05$) en over het geheel genomen significant ($p<.001$).

Vervolgens is nagegaan in hoeverre de score op de taaltoets samenhangt met de studieduur. Het blijkt dat zowel in de totale groep buitenlandse studenten met P-diploma als in de studenten bij de afzonderlijke studierichtingen een significant negatieve correlatie bestaat tussen de score op de taaltoets en de studieduur. Dit houdt in dat een hogere score op de taaltoets over het algemeen samengaat met een kortere studieduur. Over de totale groep is deze correlatie $r = -0.36$ (n=95).

6. MOGELIJKE BELEIDSMAATREGELEN EN TE VERWACHTEN EFFECTEN

Er zijn voor instellingen voor hoger onderwijs enkele beleidsmaatregelen denkbaar om het studiesucces van buitenlandse studenten te bevorderen. Deze maatregelen kunnen liggen in de sfeer van selectie en van vroegtijdige signalering in combinatie met begeleiding van potentiële uitvallers. Elk van deze maatregelen zal op grond van de resultaten van het hier beschreven onderzoek worden besproken.

Selectie van studenten

Verscheidene buitenlandse studenten die zich bij een Nederlandse instelling voor hoger onderwijs melden om een studie te gaan volgen, worden niet toegelaten. Uit cijfers van de VU te Amsterdam blijkt dat de helft van de studenten die zich aanmeldt niet wordt toegelaten tot het voorbereidend jaar op de studie (Smit, 1994). Tijdens het voorbereidend jaar valt ook nog eens eenderde van de studenten af. Deze proportie ligt voor de TU Delft in dezelfde orde van grootte (Snippe, 1994). Om te voorkomen dat veel studenten worden teleurgesteld in hun wens om in Nederland te gaan studeren, lijkt selectie in het land van herkomst een geschikt middel. Een gedegen voorlichting over wat hen hier te wachten staat, heeft al een dergelijk selecterend

Tabel 6. Gemiddelde score op taaltoets, naar resultaat bij het P-examen.

	P gehaald			geen P gehaald		
studierichting:	n	gemidd.	(s.d.)	n	gemidd.	(s.d.)
Elektrotechniek	49	8.0	(.7)	33	7.6	(.7)
Informatica	21	8.2	(.7)	29	7.7	(.7)
Werktuigbouwkunde	25	8.0	(.7)	8	7.6	(.5)
totaal	95	8.1	(.7)	70	7.7	(.7)

effect. Het zal mensen met een ontoereikende vooropleiding ontmoedigen zich toch aan te melden bij een hoger onderwijsinstelling in Nederland. Daarnaast is het mogelijk om in het land van herkomst studenten te selecteren voor een studie in het buitenland. Van de Indonesische studenten die bij de studierichtingen Elektrotechniek en Werktuigbouwkunde studeren, is bekend dat ze in eigen land een selectieproces hebben doorlopen en vervolgens een speciaal voorbereidingsprogramma hebben gevolgd, voordat ze naar Nederland kwamen. In dit programma hebben ze onderwijs gevolgd in exacte vakken en psychologie. De studenten die aan het voorbereidingsprogramma konden deelnemen, werden mede geselecteerd aan de hand van de resultaten op een intelligentietest. Ook kregen zij al les in de Nederlandse taal voordat ze naar Nederland kwamen. Er is nagegaan of deze geselecteerde studenten van de overige niet-geselecteerde buitenlandse studenten verschillen qua studiesucces. Het slaagpercentage van de geselecteerde studenten (n=62) bedraagt 71% en dat van de totale groep buitenlandse studenten (n=108) is 48%. Dit is een significant verschil. Ook de gemiddelde studieduur die nodig is om het P-examen te behalen, is voor de geselecteerde studenten korter dan voor de overige buitenlanders. Het verschil is echter niet significant. Hieruit blijkt evenwel dat een vorm van selectie effect kan sorteren. Het is alleen de vraag op welke factoren de selectie het beste kan plaatsvinden.

Een mogelijk selectiecriterium is de beheersing van de Nederlandse taal. De gemiddelde score op de toelatingstoets voor de Nederlandse taal is 7.9 voor de totale groep buitenlandse studenten in het onderzoek. Er is nagegaan of studenten met een score op de taaltoets hoger of gelijk aan deze gemiddelde score qua studiesucces verschillen van de studenten met een score op de taaltoets lager dan het gemiddelde. In tabel 7 is van deze twee groepen studenten het P-percentage weergegeven. Uit de tabel komt naar voren dat de slaagpercentages voor de studenten met een score op de taaltoets die hoger of gelijk is aan het gemiddelde, hoger zijn dan voor de studenten die een score op de taaltoets lager dan het gemiddelde hebben. De verschillen in P-rendement zijn voor de drie studierichtingen samen significant op .001-niveau, voor Informatica significant op .01-niveau en voor Werktuigbouwkunde op .05-niveau. Voor Elektrotechniek is er wel een verschil in P-rendement ten gunste van de groep met een hogere score op de taaltoets, maar dit verschil is niet significant.

Hieruit blijkt dat de selectie op grond van de beheersing van de Nederlandse taal verder zou kunnen worden aangescherpt.

Vroegtijdige signalering van uitvallers en begeleiding

Een andere mogelijke beleidsmaatregel is om in een vroeg stadium studenten te identificeren die dreigen uit te vallen. Vroegtijdige signalering kan wellicht uitval voorkomen door extra begeleiding of onderwijs in te zetten. Tevens kan de omzwaai naar een andere studierichting of naar een andere instelling voor hoger onderwijs worden ondersteund. Hiertoe is het nodig om te weten in hoeverre het mogelijk is in een vroeg stadium te voorspellen wat de kans op studiesucces zal zijn.

Van de buitenlandse studenten aan de drie faculteiten zijn de studiepunten bekend, na 6 maanden en na 12 maanden. Deze zijn in tabel 8 opgenomen.

Tabel 7. Percentage buitenlandse studenten dat het P-diploma heeft behaald, naar score op de taaltoets.

	taalscore >= 7.9		taalscore < 7.9	
studierichting:	n	percentage P	n	percentage P
Elektrotechniek	39	69%	43	51%
Informatica	26	62%	24	21%
Werktuigbouwkunde	18	94%	15	53%
totaal	83	72%	82	43%

Tabel 8. Gemiddeld aantal studiepunten van buitenlandse studenten na 6 en na 12 maanden, met tussen haakjes de standaarddeviatie.

studierichting:	n	aantal studiepunten na 6 maanden		aantal studiepunten na 12 maanden	
Elektrotechniek	83	5.1	(6.4)	16.7	(14.0)
Informatica	52	4.9	(5.1)	11.2	(10.1)
Werktuigbouwkunde	35	3.2	(4.0)	16.6	(11.8)
totaal	170	4.7	(5.7)	15.0	(12.7)

Aangezien het totaal aantal studiepunten van het propaedeuseprogramma voor Informatica 41.5 was, terwijl dit voor Elektrotechniek en Werktuigbouwkunde 42 was, zijn de aantallen studiepunten voor Informatica telkens met 42/41.5 vermenigvuldigd. Op die manier werd een directe vergelijking tussen de studierichtingen mogelijk. Het aantal studiepunten van de buitenlandse studenten bij Werktuigbouwkunde na 6 maanden is vergeleken met de andere twee studierichtingen vrij laag, terwijl dit verschil na een jaar verdwenen is. Het is mogelijk dat de verdeling van de studielast over het jaar -en dus ook de verdeling van de te behalen studiepunten- bij Werktuigbouwkunde verschillend ligt van de andere twee studierichtingen. De correlatie tussen het aantal studiepunten na 6 maanden en het aantal studiepunten na 12 maanden is .83. Het zal echter duidelijk zijn dat deze meetpunten onderling niet onafhankelijk van elkaar zijn.

Op grond van de relatie tussen de studiestand na een half jaar en de studiestand na een jaar is het wellicht mogelijk om uitgaande van de studiestand na een half jaar voorspellingen te doen over het aantal studiepunten na een jaar. Op die manier zou in een betrekkelijk vroeg stadium duidelijk kunnen zijn welke studenten als kansrijk of als minder kansrijk zijn te beschouwen ten aanzien van het behalen van het P-diploma.

Als bijvoorbeeld wordt uitgegaan van de norm die het Ministerie in het kader van de tempobeurs hanteert, dan betekent dit het volgende voor buitenlandse studenten. Vanaf het cursusjaar '93/'94 dienen studenten na een jaar 25% van het totaal aantal te behalen studiepunten in de propaedeusefase te hebben behaald om in aanmerking te komen voor studiefinanciering. Dit houdt in dat studenten na een jaar 10.5 studiepunten moeten hebben behaald. De regressievergelijking van het gemiddelde aantal studiepunten na 6 maanden op het aantal studiepunten na 12 maanden is $x = (.37 * y) - .89$. Als de y-waarde 10.5 wordt ingevuld, dan levert dit een x-waarde van 3 op, als het aantal studiepunten na 6 maanden. Het blijkt dan dat van de buitenlandse studenten die 3 of meer studiepunten hebben behaald na 6 maanden, 82% na 12 maanden aan de verplichting heeft voldaan om 10.5 of meer studiepunten te behalen. Van de studenten die na 6 maanden minder dan 3 studiepunten hebben behaald, heeft slechts 27% aan de norm van 10.5 studiepunten voldaan.

In tabel 9 worden de slaagpercentages voor het P-examen voor de totale groep buitenlandse studenten weergegeven, uitgaand van het aantal studiepunten na 6 en na 12 maanden.

Uit de tabel wordt duidelijk dat de slaagpercentages in de subgroepen met veel studiepunten significant hoger liggen dan in de subgroepen met weinig studiepunten ($p<.001$).

Veel studiepunten na een half jaar of een jaar hangen samen met een goed eindresultaat van de P-fase. Ook blijkt dat een hoger aantal studiepunten na 6 of 12 maanden samen gaat met een kortere studietijd in de P-fase, voor hen die het P-diploma behalen. De correlatie tussen het aantal studiepunten na 6 maanden en de totale studietijd in de P-fase is $r=-.56$ ($p<.001$) en de correlatie tussen het aantal studiepunten na 12 maanden en de totale studietijd in de P-fase is $r=-.76$ ($p<.001$).

Deze cijfers geven aan dat het mogelijk is om in een betrekkelijk vroeg stadium te indiceren welke studenten dreigen uit te vallen.

Tabel 9. Slaagpercentages voor P-examen en gemiddelde studieduur tot P-examen van buitenlandse studenten, naar aantallen studiepunten na 6 en na 12 maanden.

		n	percentage P	gemidd. studieduur tot P in maanden (s.d.)
na 6 maanden:	>= 3 studiepunten	84	75%	22 (8)
	< 3 studiepunten	86	38%	32 (11)
na 12 maanden:	>= 10.5 studiepunten	92	84%	23 (8)
	< 10.5 studiepunten	78	24%	36 (12)

7. CONCLUSIE EN DISCUSSIE

In dit onderzoek naar het studiesucces van buitenlandse studenten bij de studierichtingen Elektrotechniek, Informatica en Werktuigbouwkunde hebben twee onderzoeksvragen centraal gestaan. In de eerste plaats is nagegaan wat het studiesucces van buitenlandse studenten is in vergelijking met de totale groep studenten die in dezelfde periode bij de betreffende studierichting stond ingeschreven. Als tweede is de vraag onderzocht welke factoren een rol spelen bij de verklaring van verschillen in studiesucces van buitenlandse studenten.

Met betrekking tot de eerste onderzoeksvraag is gebleken dat buitenlandse studenten niet verschillen van de totale groep studenten wat betreft het slaagpercentage voor het propaedeuse-examen, met uitzondering van de studierichting Informatica. Bij deze richting zijn er significant minder buitenlandse studenten die voor hun P-examen slagen dan in de totale groep. Buitenlandse studenten hebben in vergelijking met de totale groep, wel meer tijd nodig om hun P-examen te behalen.

Een verklaring voor de betrekkelijk slechte studieresultaten van de buitenlanders bij Informatica kan op grond van de hier verzamelde en gepresenteerde gegevens niet worden gevonden in het niveau van taalbeheersing. Er was geen sprake van significante verschillen in taalbeheersing tussen buitenlandse studenten van de diverse studierichtingen onderling. De range in de scores op de taaltoets is echter niet groot aangezien alleen voldoenden zijn opgenomen. Een gevolg hiervan is dat er minder variantie in de scores aanwezig is en dat derhalve de verklaringskracht van de factor taalbeheersing afneemt. Wellicht zullen met een gevoeliger toetsinstrument voor taalbeheersing wel verschillen tussen studenten van de drie studierichtingen naar voren komen. Een andere verklaring kan zijn gelegen in de kennis van de exacte vakken. Het is mogelijk dat het beheersingsniveau van wiskunde en natuurkunde bij de buitenlandse studenten die zich bij Informatica aanmelden, lager is dan van de studenten die zich bij de andere twee studierichtingen hebben aangemeld. Hierover zijn echter geen gegevens beschikbaar.

Met betrekking tot de vraag naar verklarende factoren voor studiesucces van buitenlandse studenten is uit het onderzoek gebleken dat het land van herkomst niet leidt tot significante verschillen in studiesucces. Indonesische studenten wijken niet af van de andere buitenlandse studenten voor wat betreft hun P-rendement, noch ten aanzien van de tijd die nodig is om het P-examen te behalen. De factor sexe leidt tot verschillen in die zin dat vrouwen een lager P-rendement hebben dan mannen, maar dit verschil is niet significant. Ze hebben ook meer tijd nodig om het P-examen te behalen dan de mannen, maar dat verschil is alleen bij Elektrotechniek significant. Mannen hebben hier gemiddeld binnen de twee jaar hun P-examen gehaald, terwijl vrouwen er gemiddeld bijna $2^1/2$ jaar over doen. Op basis van de beschikbare gegevens is het niet mogelijk hiervoor een verklaring te geven.

Ook de leeftijd blijkt een factor die van belang is. Buitenlandse studenten die hun P-examen hebben behaald, zijn gemiddeld genomen een jaar jonger dan de studenten die hun P-examen niet hebben behaald. Bij de faculteit Elektrotechniek is dit zelfs anderhalf jaar. Dit komt overeen met

het gegeven dat de buitenlandse studenten die in het voorbereidingstraject op de studie uitvallen, ook gemiddeld ouder zijn dan de studenten die het traject succesvol doorlopen (Snippe, 1994). Tenslotte blijkt ook de taalvaardigheid tot verschillen in studiesucces te leiden. Studenten die hun P-examen hebben behaald, beschikken over een grotere taalvaardigheid dan zij die hun P-examen niet hebben behaald. Ook is er een negatief verband tussen taalvaardigheid en studieduur.

Onze conclusie is derhalve dat het aantal buitenlandse studenten dat het propaedeuse-examen behaalt niet zozeer afwijkt van het gemiddeld aantal studenten dat in een vergelijkbare periode dit examen behaalt, maar dat ze er meer tijd voor nodig hebben. In het begin van hun studie ondervinden buitenlandse studenten veel vertraging.

Enige relativering bij deze conclusie is op zijn plaats, vanwege het feit dat het hier gaat om studievoortgangsgegevens van buitenlandse studenten in de propaedeuse van één (technische) universiteit. Aangezien ook gegevens over het beginniveau van deze studenten ontbreken, is enige voorzichtigheid met de conclusies op zijn plaats. Opvallend is evenwel dat de gegevens overeen komen met hetgeen Admiraal & Boogaard (1994) concluderen op grond van hun globale analyses van studievoortgangsgegevens van buitenlandse studenten bij drie andere Nederlandse universiteiten.

Met inachtneming van bovenstaande relativering zijn er op grond van de geanalyseerde gegevens twee maatregelen te overwegen ter verbetering van de studievoortgang van buitenlandse studenten, namelijk selectie en vroegtijdige signalering in combinatie met begeleiding.

Selectie, bij voorkeur in het land van herkomst, lijkt een maatregel te zijn met redelijke kans op succes. Studenten die in eigen land geselecteerd zijn voor een studie in het buitenland, behalen betere studieresultaten dan studenten die niet geselecteerd zijn. Dit blijkt uit het feit dat die Indonesische studenten die in hun eigen land een selectie- en voorbereidingsprogramma hebben doorlopen op een studie in het buitenland, betere studieprestaties behalen in vergelijking met andere, niet-geselecteerde buitenlandse studenten. Ook kan de selectie in Nederland worden verscherpt. Uit de resultaten is gebleken dat buitenlandse studenten met een hogere taalbeheersingsscore dan gemiddeld, betere studieresultaten behalen in vergelijking met studenten die een taalbeheersingsscore onder het gemiddelde hebben. Een verscherping van de selectie op de beheersing van de Nederlandse taal kan dan ook leiden tot verbetering van het studiesucces van buitenlandse studenten. Een verbetering van voorlichting in het land van herkomst kan ook bijdragen aan de selectie. Het zou kunnen voorkomen dat veel buitenlandse studenten vergeefs naar Nederland komen voor een studie. In de voorlichting zou vooral nadruk moeten worden gelegd op de hoeveelheid voorkennis die wordt vereist voor een studie in Nederland, de zwaarte van de studie en de taal.

Vroegtijdige signalering in combinatie met begeleiding lijkt ook een goede mogelijkheid. Uit de resultaten is gebleken dat op grond van de studieresultaten van buitenlandse studenten na een half jaar redelijk is te voorspellen hoeveel studiepunten ze na een jaar zullen hebben behaald. In de loop van het eerste jaar kunnen daarom al in een vroeg stadium -als de studieresultaten na een half jaar bekend zijn- gesprekken plaatsvinden met buitenlandse studenten, waarin wordt nagegaan wat de oorzaken van eventuele studievertraging zijn en in hoeverre het zinvol is de studie door te zetten of naar iets anders om te zien. In combinatie met een gerichte begeleiding van buitenlandse studenten zou op die manier kunnen worden voorkomen dat studenten uitvallen. Een mentorsysteem specifiek gericht op buitenlandse studenten kan zorgen voor begeleiding in sociaal opzicht, maar het is ook mogelijk te denken aan extra begeleiding in de sfeer van extra onderwijsmaatregelen of het aanbieden van andere onderwijsvormen.

Tenslotte is het zowel in het belang van de buitenlandse student als in het belang van de onderwijsinstellingen dat alleen die mensen gaan deelnemen aan een studie, die een redelijke kans op succesvolle afronding van de studie hebben.

LITERATUUR

Admiraal, W.F. & Boogaard, M. (1994). Buitenlandse studenten aan de universiteiten van Eindhoven, Utrecht en Wageningen. In J. Snippe, W.M.G. Jochems en W. van den Bor (red.). *Buitenlandse studenten in het Nederlandse hoger onderwijs.* Delft: Delftse Universitaire Pers.

Ayers, J.B. & Peters, R.M. (1977). Predictive validity of the test of English as a foreign language for Asian graduate students in engineering, chemistry, or mathematics. *Educational and Psychological Measurement, 37,* 461-463.

Fraser, B.J. et al. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11,* 145-252.

Graham, J.G. (1987). English language proficiency and the prediction of academic success. *Tesol Quarterly, 3,* 505-521.

Ho, D.Y.F. & Spinks, J.A. (1985). Multivariate prediction of academic performance by Hong Kong university students. *Contemporary Educational Technology, 10,* 249-259.

Jochems, W. (1991). Effects of learning and teaching in a foreign language. *European Journal of Engineering Education, 16,* 4, pp. 309-316.

Jochems, W., Rijns, Y., Smid, H.J. & Verweij, A. (1993). *The academic progress of foreign students of computer science at Delft University of Technology.* Delft: Faculty of Technical Mathematics and Informatics, report 93-60.

Jochems, W., Smid, H.J., Snippe, J. & Verweij, A. (1994). *Studiesucces van buitenlandse studenten aan de TU Delft* (in voorbereiding).

Light, R., Xu, M. & Mossop, J. (1987). English proficiency and academic performance of international students. *Tesol Quarterly, 2,* 251-261.

Overwalle, F. van (1985). Slagen en mislukken in de eerste kandidatuur: op zoek naar oorzaken. *Tijdschrift voor Onderwijsresearch, 10,* 107-129.

Sharon, A.T. (1972). English proficiency, verbal aptitude, and foreign student success in American graduate schools. *Educational and Psychological Measurement, 32,* 425-431.

Smit, C.P. (1994). Buitenlandse studenten in het academisch voortraject. In J. Snippe, W.M.G. Jochems & W. van den Bor (red.). *Buitenlandse studenten in het Nederlandse hoger onderwijs.* Delft: Delftse Universitaire Pers.

Snippe, J. (1994). Noncognitieve kenmerken van buitenlandse studenten bij aanvang van de studie. In J. Snippe, W.M.G. Jochems & W. van den Bor (red.). *Buitenlandse studenten in het Nederlandse hoger onderwijs.* Delft: Delftse Universitaire Pers.

Snippe, J., Jochems, W.M.G. & Bor, W. van den (1994) (red.). *Buitenlandse studenten in het Nederlandse hoger onderwijs.* Delft: Delftse Universitaire Pers.

Statistisch Jaarboek 1992/1993. Delft: Technische Universiteit, 1993.

Walberg, H.J. (1986). Een synthese van onderzoek naar onderwijsproductiviteit. In W.J. Nijhof & E. Warries (red.). *De Opbrengst van Onderwijs en Opleiding.* Lisse: Swets & Zeitlinger, 15-24.

Wellens, J. (1980). De predictie van het studieresultaat in het hoger onderwijs: een terreinverkenning. Brussel: Vrije Universiteit.

Wende, M. van der, (1992). De instroom van buitenlandse studenten in het Nederlandse hoger onderwijs. *Visum Perspectief, 7.* Den Haag: Nuffic.

Manuscript ontvangen 14-7-1994
Definitieve versie ontvangen 24-10-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 2, pp. 178-190

Huiswerk maken bij de televisie of geluidsmedia*

J.W.J. Beentjes, C.M. Koolstra en T.H.A. van der Voort
R.U. Leiden, Sectie Kind en Media

ABSTRACT

Combining homework with television or audio media
A survey conducted among 1700 students from the second and fourth grade of secondary school investigated how often pupils carry out different types of homework assignments and how often these are combined with media. In addition, it was investigated how pupils divide their attention between homework assignments and background media, and how they perceive the effects of media on different types of homework assignments. The frequency with which homework and media are combined appears to depend on the type of homework (learning assignments or paper-and-pencil assignments), medium type (audio media or television) and the students' level of secondary education. According to the pupils, learning assignments are impaired when combined with background media, whereas paper-and-pencil assignments are not impaired by audio media or musictelevision.

Het effect van achtergrondmedia op de kwaliteit van het gemaakte huiswerk is nooit rechtstreeks onderzocht. Wel is experimenteel onderzoek verricht naar de invloed van achtergrondtelevisie en achtergrondgeluiden op de uitvoering van verschillende cognitieve taken. Deze studies (Armstrong, 1993; Armstrong, Boiarsky & Mares, 1991; Armstrong & Greenberg, 1990) wijzen uit dat achtergrond*televisie* interfereert met moeilijke cognitieve taken die veel aandacht vereisen. Het combineren van een complexe taak met televisie kijken lijkt de aandachtscapaciteit van mensen te boven te gaan. Daarnaast zijn er aanwijzingen dat achtergrondtelevisie interfereert met het verwerken van visueel-ruimtelijke informatie, zoals vastgesteld met een test bestaande uit geometrische analogieproblemen (Armstrong, 1993).

In het onderzoek naar de invloed van achtergrond*geluid* op cognitieve taken is onder meer gekeken naar het effect van diverse typen achtergrondgeluiden op begrijpend lezen (Martin, Wogalter & Forlano, 1988). In een serie experimenten werden de proefpersonen geïnstrueerd zich volledig op de tekst te concentreren en de achtergrondgeluiden te negeren. Ondanks deze instructie bleek een akoestische achtergrond van betekenisvolle woorden, vergeleken met betekenisloze woorden, instrumentale muziek, ruis en stilte, het begrijpend lezen te hinderen. De onderzoekers concludeerden dat men niet ontkomt aan het verwerken van betekenisvolle achtergrondgeluiden ('mandatory processing'). Daarnaast lijkt achtergrondmuziek het korte-termijngeheugen te storen. Bij het onthouden van getallen bestaande uit negen cijfers was vocale muziek hinderlijker dan instrumentale muziek; instrumentale muziek was hinderlijker dan ruis of stilte, terwijl tussen ruis en stilte geen verschil werd geconstateerd (Salamé & Baddely, 1989).

Hoewel deze experimenten doen vermoeden dat bepaalde huiswerktaken zich slecht laten combineren met televisie en geluidsmedia op de achtergrond, hebben zij een beperkte zeggingskracht voor de huiswerksituatie van (Nederlandse) scholieren. Ten eerste zijn alle experimenten uitgevoerd bij studenten die, in vergelijking met scholieren in het voortgezet onderwijs, over relatief sterk ontwikkelde zelfstudievaardigheden beschikken en zich relatief goed kunnen concentreren. Bovendien kunnen Nederlandse scholieren, in tegenstelling tot de Amerikaanse proefpersonen, de teksten van de Engelstalige songs, die ook in Nederland het leeuwedeel van de beluisterde muziek uitmaken, slechts gedeeltelijk verstaan. Ten tweede is in de experimenten

*Dit onderzoek kwam tot stand met steun van een ANGO-subsidie van SVO.
Adres: Postbus 9555, 2300 RB Leiden.

gebruik gemaakt van kortdurende taken die niet meer dan enkele minuten duren, terwijl huiswerktaken meestal meer tijd in beslag nemen. Bovendien werkten de proefpersonen meestal aan kunstmatige taken die ver van huiswerktaken afstaan. Tenslotte is ook de levensechtheid van de gebruikte achtergrondstimuli twijfelachtig. Zo zijn de experimentele taken niet uitgevoerd bij videoclips, een onder de huidige jeugd zeer populair programmatype.

Om uitspraken te kunnen doen over de invloed van achtergrondmedia op huiswerkprestaties zijn experimentele studies nodig met een hogere ecologische validiteit voor de (Nederlandse) huiswerksituatie. Een goede invulling van deze experimenten vereist specifieke informatie over de huiswerksituatie van Nederlandse scholieren, met name welke huiswerktaken het vaakst met welke media gecombineerd worden. In Nederland is weliswaar onderzoek gedaan naar het maken van huiswerk, maar daarbij is nauwelijks aandacht besteed aan het combineren van huiswerk met achtergrondmedia. Een enquête onder 3000 leerlingen uit de eerste vier leerjaren van diverse typen van het voortgezet onderwijs (Neuwahl & Van den Bogaart, 1989) wijst over ons onderwerp alleen uit dat ongeveer de helft van de leerlingen meer of minder frequent het maken van huiswerk met de radio en/of tv (die in deze enquête waren samengenomen) combineerde.

In voorbereiding op later uit te voeren experimentele studies wordt in het hier besproken enquête-onderzoek nagegaan (a) hoe vaak leerlingen verschillende typen huiswerktaken uitvoeren, (b) hoe vaak zij deze huiswerktaken met onderscheiden media en media-inhouden combineren, (c) hoe kinderen hun aandacht over huiswerk en media verdelen, en (d) hoe kinderen zelf het effect van achtergrondmedia op verschillende huiswerktypen inschatten.

Omdat het huiswerk maken pas na de basisschool een algemeen verschijnsel wordt, richt het onderzoek zich op leerlingen van het voortgezet onderwijs. In het eerste leerjaar van het voortgezet onderwijs heeft het studeergedrag thuis nog geen vaste vorm gekregen, zeker aan het begin van het schooljaar, het tijdstip waarop het hier gerapporteerde onderzoek is uitgevoerd. Het onderzoek is daarom uitgevoerd bij leerlingen van het tweede leerjaar. Daarnaast is het onderzoek uitgevoerd in het vierde leerjaar, een leerjaar waarin scholieren uit alle schooltypen ruime ervaring met huiswerk hebben opgedaan. Omdat er tussen schooltypen verschillen bestaan in de aard van het huiswerk, en mogelijk ook in de wijze waarop media ermee worden gecombineerd, zijn vier schooltypen in het onderzoek betrokken: het VBO (het vroegere LBO), het MAVO, het HAVO en het VWO.

METHODE

Constructie van de vragenlijst

Ter ontwikkeling van de in het onderzoek gebruikte vragenlijst zijn twee voorstudies uitgevoerd. In een eerste vooronderzoek zijn semigestructureerde interviews met 20 scholieren uit de tweede en vierde klas van het voortgezet onderwijs gehouden om tot een indeling van huiswerktaken, achtergrondmedia en media-inhouden te komen. In het tweede vooronderzoek is mede op basis van deze interviews een schriftelijke proefversie van de in het hoofdonderzoek te gebruiken enquête opgesteld. Om mogelijke misverstanden die de items oproepen op het spoor te komen is aan 24 leerlingen uit beide leerjaren van de vier onderscheiden schooltypen gevraagd de vragen uit deze proefenquête hardop denkend in aanwezigheid van een proefleider te beantwoorden. Op basis van de bevindingen met de proefenquête is de definitieve vragenlijst opgesteld.

Op grond van het vooronderzoek zijn in de vragenlijst twee typen huiswerk onderscheiden: *maakwerk* en *leerwerk*. Leerwerk is het huiswerk waarbij men dingen uit het hoofd moet leren of waarbij men teksten moet lezen die men op een later tijdstip in hoofdlijnen moet kunnen reproduceren. Maakwerk is huiswerk waarbij thuis opdrachten moeten worden uitgevoerd waarvan de resultaten op school getoond moeten worden, maar die men niet van buiten hoeft te kennen. Onder de noemer maakwerk valt bijvoorbeeld het oplossen van wiskundeopgaven, het beantwoorden van vragen over een gelezen tekst, het opzoeken van informatie en het maken van werkstukken.

Respondenten

Aan het onderzoek is medegewerkt door 20 scholen voor voortgezet onderwijs in de provincie Zuid-Holland. Er zijn uitsluitend scholen benaderd uit bekabelde regio's, zodat het aanbod van met name televisie (MTV, RTL4) voor alle respondenten gelijk is. Bij de selectie van scholen is gestreefd naar een evenredige vertegenwoordiging van leerlingen naar leerjaar, schooltype en sekse. Omdat de vrijetijdsbesteding van kinderen binnen en buiten de grote stad kan verschillen, zijn de scholen zo gekozen dat de helft van de leerlingen woonachtig is in een grote stad terwijl de andere helft van buiten de grote stad komt. Tevens is getracht een evenredige vertegenwoordiging van scholen naar denominatie tot stand te brengen. Om tot de gewenste steekproef van 20 scholen te komen, zijn in totaal 34 scholen benaderd, waarvan er 11 niet wilden meedoen en drie niet aan de genoemde vereisten voldeden. Per school zijn in overleg met de schoolleiding maximaal vier klassen geselecteerd. De enquête is door alle in de klas aanwezige leerlingen ingevuld. In tabel 1 is het aantal ondervraagden per leerjaar en schooltype weergegeven. Omdat op veel scholen pas in het derde leerjaar wordt besloten welke leerlingen naar het HAVO en het VWO gaan, is het niet mogelijk voor het tweede leerjaar een onderscheid te maken tussen leerlingen die het HAVO en het VWO volgen. Om die reden onderscheiden we in onze analyses slechts drie schooltypen, namelijk VBO, MAVO en HAVO/VWO.

Vergeleken met de meest recente cijfers van het CBS (1993) komt de demografische samenstelling van de geselecteerde steekproef redelijk overeen met die van de landelijke schoolbevolking in het voortgezet onderwijs. In de geselecteerde steekproef zijn leerlingen uit het tweede leerjaar (52%) iets talrijker dan in leerjaar 4 (48%). Ook landelijk telt het tweede leerjaar (51%) iets meer leerlingen dan het vierde leerjaar (49%). De in tabel 1 vermelde verdeling naar schooltypen komt eveneens goeddeels overeen met het landelijke beeld (VBO: 22%, MAVO: 34%, HAVO/VWO: 44%), zij het dat de MAVO-leerlingen in onze steekproef (37,5%) enigszins oververtegenwoordigd zijn. Conform het landelijk beeld (meisjes: 52%) komen ook in onze steekproef meisjes (54%) iets frequenter voor dan jongens. Ook de beoogde fifty-fifty-verdeling van stads- en plattelandskinderen ziet men in de geselecteerde steekproef terug (respectievelijk 52% en 48%). In vergelijking met de landelijke verdeling van PC-, RK- en Openbaar-neutrale scholen (respectievelijk 25%, 32% en 43%) zijn in de geselecteerde steekproef de PC-scholen (31%) enigszins oververtegenwoordigd en de openbare scholen (39%) enigszins ondervertegenwoordigd.

Procedure

De scholen zijn in de periode oktober-november 1993 bezocht door een of meer onderzoekers die de vragenlijsten introduceerden en vervolgens door de leerlingen zelfstandig lieten invullen.

RESULTATEN

Om na te gaan in hoeverre de in dit onderzoek onderscheiden achtergrondvariabelen (schooltype, leerjaar, sekse, woonplaats binnen of buiten een grote stad, denominatie van de school) een rol spelen in het huiswerkgedrag is een aantal multiple regressieanalyses uitgevoerd waarin

Tabel 1. Samenstelling van de steekproef naar leerjaar en schooltype (N = 1700).

Leerjaar 2				Leerjaar 4			
VBO	MAVO	HAVO/ VWO	**totaal**	VBO	MAVO	HAVO/ VWO	**totaal**
198	335	359	**892**	163	297	348	**808**
22%	38%	40%	**100%**	20%	37%	43%	**100%**

is vastgesteld welke achtergrondvariabelen een onafhankelijke bijdrage leveren aan de voorspelling van de volgende aspecten van huiswerkgedrag: tijd besteed aan huiswerk, de frequentie van het gebruik van achtergrondmedia tijdens het huiswerk en de gepercipieerde effecten van achtergrondmedia op de huiswerkprestaties. Slechts een van de achtergrondvariabelen, schooltype, bleek een rol te spelen bij de voorspelling van elk van de onderscheiden afhankelijke variabelen. De overige achtergrondvariabelen leverden slechts incidenteel een onafhankelijke bijdrage aan de voorspelling van het huiswerkgedrag. Mede om de presentatie van de resultaten overzichtelijk te houden is daarom in de hieronder te bespreken variantieanalyses alleen de variabele 'schooltype' als factor opgenomen. Indien andere achtergrondvariabelen mede van invloed zijn, is dit aangegeven door in de tekst de β-waarden zoals gevonden in de genoemde multiple regressieanalyses op te nemen, althans wanneer er sprake is van een significante bijdrage ($p < 0{,}01$) aan het huiswerkgedrag.

Omdat voorstelbaar is dat verschillen op achtergrondvariabelen feitelijk zijn toe te schrijven aan verschillen tussen scholen, is middels een zogenaamde multiniveau-analyse (Goldstein, 1987) nagegaan of individuele verschillen in huiswerkgedrag (niveau 1) zijn veroorzaakt door verschillen tussen scholen (niveau 2). Hiertoe is het huiswerkgedrag voorspeld met de achtergrondvariabelen als vaste effecten en individu en school als random-effecten. Uit de resultaten van de multiniveau-analyse bleek dat er in geen van de hieronder gerapporteerde analyses sprake is van een schooleffect.

Tijd besteed aan huiswerk

De tijd die aan huiswerk wordt besteed, is vastgesteld door de respondenten te vragen voor elke afzonderlijke dag van de week te schatten hoeveel uren en kwartieren zij aan respectievelijk maakwerk en leerwerk besteden. Deze zevenvoudige schatting (voor elke dag van de week)

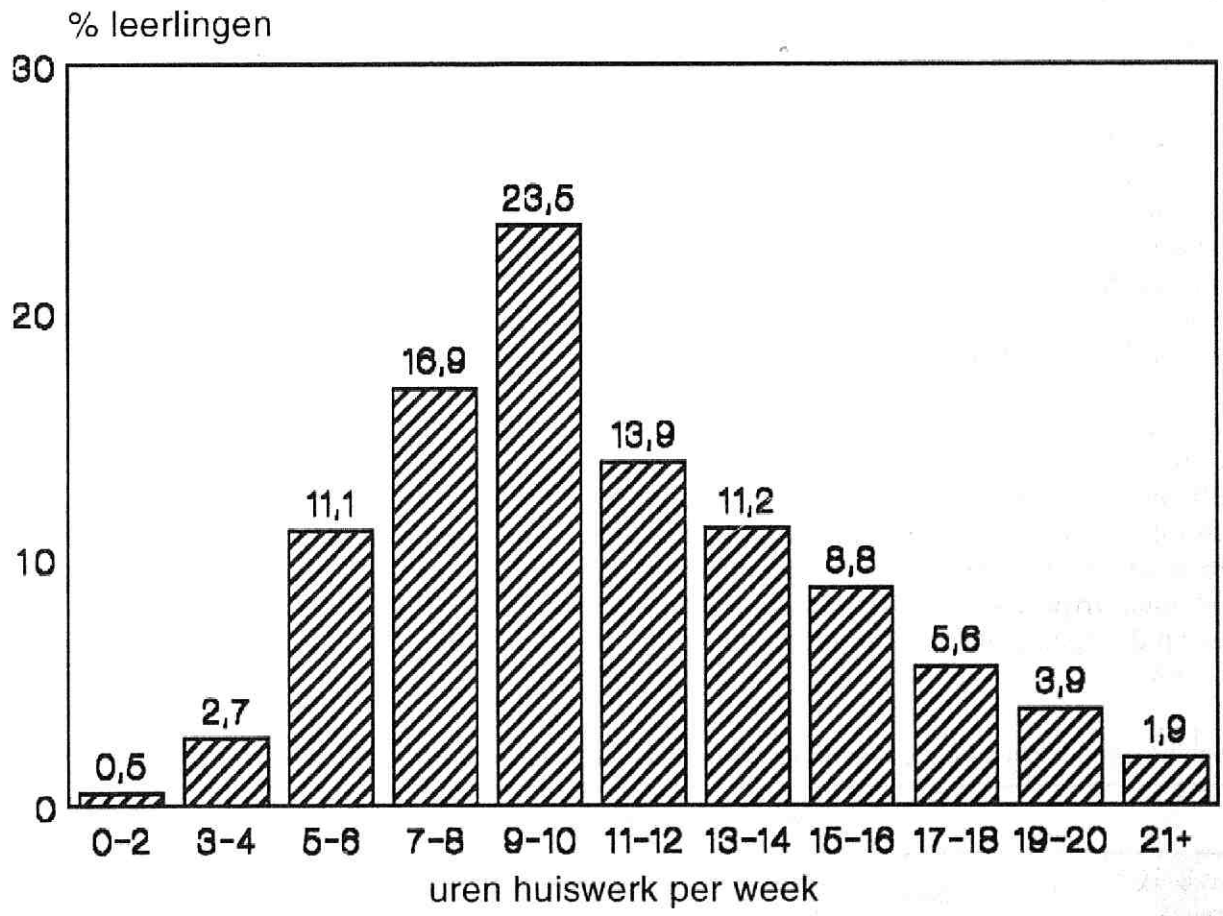


Fig. 1. Procentuele frequentieverdeling van het aantal uren besteed aan huiswerk.

achten wij betrouwbaarder dan de eenmalige schatting van de gemiddelde tijd per dag zoals die bijvoorbeeld in het huiswerkonderzoek van Neuwahl en Van den Bogaart (1989) is gehanteerd, omdat de tijd die aan huiswerk wordt besteed met de dag (en het lesrooster) kan wisselen.

Er bestaan enorme verschillen in de hoeveelheid tijd die de respondenten per week voor het maken van huiswerk uittrekken (zie Figuur 1). Ongeveer 2% van de leerlingen zegt meer dan 20 uur per week aan huiswerk te besteden, terwijl 3% minder dan 4 uur per week voor het huiswerk uittrekt. Ruim de helft van de leerlingen besteedt tussen de 6 en 12 uur per week aan het huiswerk. Het gemiddelde over alle leerlingen is ongeveer 11 uur per week.

Blijkens een 2x3-variantieanalyse met als binnen-proefpersonen-factor Huiswerktype (maakwerk en leerwerk) en als tussen-proefpersonen-factor Schooltype (VBO, MAVO, en HAVO/VWO) is er een hoofdeffect van zowel huiswerktype als schooltype. Zoals te zien in tabel 2 wordt in het algemeen meer tijd besteed aan leerwerk (6 uur per week) dan aan maakwerk (5 uur per week), $F(1, 1692) = 85{,}50$, $p < 0{,}001$. Daarnaast neemt de tijd besteed aan huiswerk toe naarmate het schooltype hoger is, $F(2,1692) = 13{,}89$, $p < 0{,}001$. Er is echter ook een significante interactie tussen huiswerktype en schooltype, $F(2, 1992) = 21{,}04$, $p < 0{,}001$, hetgeen betekent dat beide verbanden niet onverkort opgaan. De regel dat aan leerwerk meer tijd wordt besteed dan aan maakwerk blijkt alleen voor het (meer theoretisch ingestelde) algemeen vormend onderwijs op te gaan (MAVO, HAVO/VWO) en niet voor het meer praktisch georiënteerde VBO. De regel dat de tijd besteed aan huiswerk toeneemt naarmate het schooltype hoger is, blijkt volgens een post-hoc-vergelijking met behulp van de Scheffé-test ($p < 0{,}01$) alleen op te gaan voor het leerwerk. Wat betreft de tijd besteed aan maakwerk bestaat er geen significant verschil tussen de drie schooltypen, terwijl VBO-leerlingen voor leerwerk (M = 4,8 uur) significant minder tijd uittrekken dan leerlingen van MAVO en HAVO/VWO (M = 6,2 uur).

Blijkens multiple regressieanalyse vertonen naast schooltype twee andere achtergrondvariabelen, sekse en denominatie van de school, samenhang met de tijd die aan huiswerk en met name aan leerwerk wordt besteed. Meisjes besteden meer tijd aan leerwerk dan jongens (β = 0,14, $p < 0{,}01$) en op PC-scholen wordt meer tijd aan leerwerk besteed dan op RK- en openbare scholen ($\beta = 0{,}08$, $p < 0{,}01$).

In de enquête is nagevraagd hoe vaak de respondenten hun huiswerk op de eigen kamer dan wel in de huiskamer maken (altijd/vaak/soms/nooit). Dit gebeurt het meest op de eigen kamer. Ruim 80% van de leerlingen maakt het huiswerk vaak of altijd op de eigen kamer, terwijl 5% nooit huiswerk op de eigen kamer doet. De huiskamer is zelden het toneel van huiswerk makende kinderen: nog geen 15% doet het huiswerk vaak of altijd in de huiskamer, terwijl 44% van de respondenten nooit huiswerk in de huiskamer maakt.

De combinatie van huiswerk met achtergrondmedia

Aanwezige apparatuur

De eigen kamer van de respondenten is rijkelijk voorzien van geluidsmedia. Bijna iedereen heeft een radio en een cd- of cassettespeler op de eigen kamer en meer dan driekwart heeft op deze lokatie een walkman (zie tabel 3). Daarnaast blijkt opvallend genoeg niet minder dan de helft over een televisietoestel op de eigen kamer te beschikken. De genoemde media zijn in de huiskamer vrijwel alom aanwezig, dit met uitzondering van de walkman die blijkbaar bij voorkeur op de eigen kamer wordt gebruikt.

Tabel 2. Gemiddeld aantal uren per week besteed aan maakwerk en leerwerk per schooltype.

	VBO	MAVO	HAVO/VWO	**totaal**
maakwerk	4,89	4,91	5,17	**5,01**
leerwerk	4,81	6,15	6,34	**5,95**
totaal	**9,70**	**11,06**	**11,51**	**10,96**

Tabel 3. Percentage leerlingen dat beschikt over audiovisuele apparatuur op de eigen kamer en in de huiskamer.

Apparatuur	Eigen kamer	Huiskamer
radio	92,1	94,6
cd- of cassettespeler	89,6	94,1
walkman	76,6	22,8
televisie	51,1	98,1

De aanwezigheid van media betekent uiteraard niet dat deze ook tijdens het huiswerk aanstaan. Op de vraag of de beschikbare media tijdens het huiswerk 'wel eens' aan staan, antwoordt 18% van de respondenten de media op de eigen kamer tijdens het huiswerk nooit aan te doen. Bij het maken van huiswerk in de huiskamer – hetgeen, zie boven, zelden voorkomt – blijven media bij 32% van de respondenten ongebruikt.

Van degenen die op de eigen kamer over geluidsmedia beschikken, combineert een ruime meerderheid het maken van huiswerk wel eens met het luisteren naar cd- of cassettespeler (80%) of de radio (70%), terwijl slechts 19% het huiswerk wel eens maakt met een walkman op. Van de respondenten met een televisie op de eigen kamer, zet 56% het toestel tijdens het huiswerk wel eens aan.

In de huiskamer komt – evenals op de eigen kamer – de combinatie huiswerk/walkman zelden voor. Slechts 15% van de respondenten die in de huiskamer over een walkman beschikken, gebruikt dit medium tijdens het huiswerk. Ongeveer de helft van de respondenten die in de huiskamer over een radio (52%) of een cd- of cassettespeler (49%) beschikken, hebben deze media tijdens het huiswerk wel eens aan; voor de televisie bedraagt dit percentage 53.

De frequentie van het gebruik van achtergrondmedia

De respondenten is gevraagd aan te geven hoe frequent zij hun leerwerk en hun maakwerk combineren met televisie en geluidsmedia: 'altijd' (100% van de aan huiswerk bestede tijd), 'heel vaak' (75% van de tijd), 'vaak' (50%), 'soms' (25%), of 'nooit' (0%). Blijkens een 4x3-varianticanalyse met als binnen-proefpersonen-factor Combinatie van huiswerktype en medium (maakwerk/geluidsmedia, leerwerk/geluidsmedia, maakwerk/televisie en leerwerk/televisie) en als tussen-proefpersonen-factor Schooltype (VBO, MAVO, en HAVO/VWO) is sprake van een hoofdeffect van zowel huiswerk/medium-combinatie als schooltype (zie tabel 4). Het hoofdeffect van huiswerk/medium-combinatie houdt in dat de frequentie waarmee huiswerk wordt

Tabel 4. Gemiddelde frequentie waarmee maak- en leerwerk worden gecombineerd met geluidsmedia of televisie per schooltype.

Combinatie	VBO	MAVO	HAVO/VWO	**totaal**
maakwerk/ geluidsmedia	3,74	3,50	3,15	**3,41**
leerwerk/ geluidsmedia	2,48	2,16	1,88	**2,11**
maakwerk/ televisie	2,12	1,93	1,68	**1,87**
leerwerk/ televisie	1,64	1,46	1,30	**1,43**
totaal	**2,50**	**2,26**	**2,00**	**2,21**

Noot: De schaalwaarden zijn: nooit (1), soms (2), vaak (3), heel vaak (4) en altijd (5)

gecombineerd met achtergrondmedia afhangt van zowel het type medium als het type huiswerk, $F(3, 5082) = 1207{,}56$, $p < 0{,}001$). Post-hoc-vergelijking met behulp van de Scheffé-test ($p < 0{,}05$) laat zien dat de frequenties waarmee de vier huiswerk/medium-combinaties worden gebruikt alle significant van elkaar verschillen. De meest frequent gebruikte combinatie is de combinatie maakwerk/geluidsmedia, gevolgd door leerwerk/geluidsmedia en maakwerk/televisie; de minst gebruikte combinatie is leerwerk/televisie. Het hoofdeffect van schooltype houdt in dat het huiswerk vaker met media wordt gecombineerd naarmate het schooltype lager is, $F(2,1694) = 40{,}84$, $p < 0{,}001$). Post-hoc-vergelijking met behulp van de Scheffé-test ($p < 0{,}05$) wijst echter uit dat alleen het verschil tussen HAVO/VWO en VBO significant is.

Multiple regressieanalyse wijst uit dat naast schooltype ook het leerjaar (2 of 4) en de woonplaats (binnen of buiten de grote stad) samenhangen met de frequentie waarmee huiswerk met achtergrondmedia wordt gecombineerd. In leerjaar 4 staan tijdens het huiswerk vaker geluidsmedia aan dan in leerjaar 2, zowel bij maakwerk ($\beta = 0{,}13$, $p < 0{,}01$) als bij leerwerk ($\beta = 0{,}08$, $p < 0{,}01$), terwijl televisie in beide leerjaren even frequent met huiswerk wordt gecombineerd. Leerlingen die buiten de grote stad wonen, hebben tijdens hun huiswerk vaker de televisie aan staan dan leerlingen in een grote stad, zowel bij leerwerk ($\beta = 0{,}10$, $p < 0{,}01$) als bij maakwerk ($\beta = 0{,}07$, $p < 0{,}01$), terwijl er geen samenhang is tussen woonplaats en de combinatie van huiswerk met geluidsmedia.

Media-inhouden waarmee huiswerk wordt gecombineerd

Bijna alle respondenten (99%) die het maken van huiswerk met het luisteren naar *geluidsmedia* combineren, gebruiken geluidsmedia om naar achtergrondmuziek te luisteren. Hoewel de meeste leerlingen die tijdens hun huiswerk muziek aan hebben meerdere muzieksoorten noemen waarnaar zij tijdens het huiswerk luisteren (o.a. heavy metal, rap, klassiek en bubbling), zijn er drie duidelijk favoriet: pop (70%), house (55%) en rock (25%).

Als leerlingen muziek willen horen tijdens hun huiswerk stemmen zij af op een radiozender of draaien zij een zelf gekozen cassette of cd af. Van de leerlingen die bij het huiswerk wel eens de radio aan hebben, stemmen de meesten (65%) steeds op dezelfde zender af. Deze zendertrouwe luisteraars stemmen meestal af op een commerciële muziekzender (42%) zoals Sky radio, RTL radio, of Radio 538, gevolgd door Radio 3 (27%) en de lokale radiozenders (25%).

Het percentage leerlingen dat tijdens het huiswerk wel eens de *televisie* aan heeft, blijft weliswaar ver achter bij de groep leerlingen die wel eens met geluidsmedia aan huiswerk maakt, maar bedraagt toch nog ongeveer 50%. De meeste van deze leerlingen (76%) hebben tijdens het huiswerk muziekprogramma's op de televisie aanstaan, maar 53% van de leerlingen noemt ook dramaseries. Sportprogramma's en praatprogramma's komen op de derde plaats. Deze programma's worden elk door 23% van de respondenten met huiswerk gecombineerd.

Op de vraag welke concrete programma's tijdens het huiswerk 'meestal' aanstaan, worden de programma's van de muziekzender MTV, die voornamelijk uit videoclips bestaan, het vaakst genoemd. Bij ongeveer 70% van de leerlingen die wel eens huiswerk bij de televisie maken, staat het toestel meestal afgestemd op MTV. Daarnaast maakt 30 à 40% huiswerk met soap- en dramaseries als achtergrond, met name *Beverly Hills 90210*, *Vrienden voor het Leven*, *Goede Tijden Slechte Tijden* en *The Bold and the Beautiful*.

Van de leerlingen die wel eens huiswerk maken bij de televisie stemt ongeveer 40% steeds op dezelfde zender af. Deze zendertrouwe kijkers stemmen in meerderheid af op RTL4 (47%) of MTV (45%). Daarnaast maakt een kleine groep (3%) huiswerk bij Eurosport. De overige leerlingen hebben nu eens de ene dan weer de andere zender aan.

Verdeling van aandacht

De wijze waarop leerlingen hun aandacht over huiswerk en achtergrondmedia verdelen is onderzocht door hun te vragen in hoeverre een reeks uitspraken op hen van toepassing is (altijd/vaak/soms/nooit). In de hieronder gerapporteerde percentages zijn de respondenten die op een uitspraak 'vaak' of 'altijd' antwoorden samengenomen. Wanneer leerlingen huiswerk combineren met muziek via geluidsmedia wordt weliswaar het huiswerk gezien als de primaire taak,

maar de aandacht voor de achtergrondmuziek blijkt sterk te kunnen variëren. De meeste leerlingen (67%) herkennen zich in de uitspraak: 'Als ik mijn huiswerk doe met muziek aan, hoor ik de muziek op de achtergrond maar mijn hoofd is bij het huiswerk.' Bij ruim eenderde (37%) van de leerlingen eist de achtergrondmuziek echter een groter deel van de aandacht op, omdat zij aangeven tijdens het huiswerk maken 'tegelijkertijd goed naar de muziek te luisteren.' Voor ruim eenvijfde van de leerlingen is het luisteren naar achtergrondmuziek wisselend een secundaire en primaire taak, omdat zij 'met hun hoofd nu eens bij het huiswerk en dan weer bij de muziek zitten' (27%), of frequent 'met het huiswerk stoppen om naar de muziek te luisteren' (21%). Bij een minderheid (17%) verdwijnt de muziek echter doorgaans geheel uit de aandacht doordat men 'vergeet dat er muziek aanstaat.'

Er wordt wel geopperd dat de televisie tijdens het huiswerk maken als een soort radio functioneert. Dit is voor een grote minderheid (41%) van de respondenten inderdaad het geval: men heeft 'het hoofd bij het huiswerk maar hoort de televisie op de achtergrond.' De meeste leerlingen die tijdens het huiswerk maken de televisie aan hebben kijken echter wel degelijk naar de televisie: 59% van de leerlingen 'volgt tijdens het huiswerk tegelijkertijd het televisieprogramma door af en toe even te kijken.' Het patroon dat de aandacht wisselend uitgaat naar het huiswerk en de televisie komt vaker voor dan bij het huiswerk maken met muziek. Ongeveer tweevijfde van de leerlingen zegt dat zij vaak 'stoppen met hun huiswerk om naar de televisie te kijken' (45%) of 'met hun hoofd nu eens bij het huiswerk en dan weer bij het televisieprogramma zitten' (38%). Maar evenals bij de geluidsmedia kan ook de televisie tijdens het huiswerk maken geheel uit de aandacht verdwijnen: 15% van de leerlingen 'vergeet vaak dat men de televisie heeft aanstaan.'

Gepercipieerde effecten van het combineren van huiswerk met achtergrondmedia

Om na te gaan hoe de leerlingen zelf denken over de effecten van het combineren van huiswerk met achtergrondmedia, is de respondenten gevraagd te reageren op een negental stellingen over de invloed die het gebruik van achtergrondmedia op het maken van huiswerk kan hebben door te kiezen uit een positief, een negatief en een neutraal alternatief. Bijvoorbeeld: 'Televisie bij maakwerk zorgt ervoor: dat het maakwerk sneller af is/dat het maakwerk minder snel af is/maakt niet uit' (zie Appendix). In de enquête zijn deze negen stellingen voor elke combinatie van beide typen huiswerk (maakwerk en leerwerk) en beide achtergrondmedia (televisie en geluidsmedia) herhaald. De vier resulterende schalen voor de combinaties maakwerk/geluidsmedia (Cronbachs alfa = 0,88), leerwerk/geluidsmedia (alfa = 0,91), maakwerk/televisie (alfa = 0,86) en leerwerk/televisie (alfa = 0,84) hebben een bevredigende betrouwbaarheid.

Blijkens een 4x3-variantieanalyse met als binnen-proefpersonen-factor Combinatie van huiswerktype en medium (maakwerk/geluidsmedia, leerwerk/geluidsmedia, maakwerk/televisie en

Tabel 5. Gemiddelde gepercipieerde effecten van achtergrondmedia op huiswerk.

Combinatie	VBO	MAVO	HAVO/VWO	**totaal**
maakwerk/ geluidsmedia	+0,18	+0,09	0,00	**+0,07**
maakwerk/ televisie	–0,26	–0,35	–0,45	**–0,37**
leerwerk/ geluidsmedia	–0,24	–0,38	–0,48	**–0,39**
leerwerk/ televisie	-0,45	–0,52	–0,62	**–0,55**
totaal	**–0,19**	**–0,29**	**–0,39**	**–0,31**

Noot: De scores variëren van -1 (combinatie heeft sterk negatief effect) tot +1 (combinatie heeft sterk positief effect)

leerwerk/televisie) en als tussen-proefpersonen-factor Schooltype (VBO, MAVO, en HAVO/VWO) zijn er hoofdeffecten van zowel huiswerk/medium-combinatie als schooltype. Zoals tabel 5 laat zien, houdt het hoofdeffect van schooltype in dat leerlingen in hogere schooltypen de effecten van achtergrondmedia op huiswerk negatiever inschatten dan leerlingen in lagere schooltypen, $F(2,1694) = 38,43$, $p < 0,001$. Post-hoc-vergelijking met behulp van de Scheffé-test ($p < 0,05$) wijst uit dat alleen het verschil tussen HAVO/VWO en VBO significant is. Het hoofdeffect van huiswerk/medium-combinatie, $F(3, 4782) = 983,17$, $p < 0,001$, houdt in dat de gepercipieerde effecten van achtergrondmedia op huiswerk afhangen van het type medium en het soort huiswerk. Gemiddeld genomen denken de leerlingen dat maakwerk niet beter maar ook niet slechter gaat met geluidsmedia op de achtergrond, terwijl de invloed van geluidsmedia op leerwerk en het effect van televisie op zowel maakwerk als leerwerk negatief worden ingeschat. Post-hoc-vergelijking met behulp van de Scheffé-test ($p < 0,05$) wijst uit dat de respondenten de invloed van geluidsmedia op maakwerk significant minder negatief inschatten dan het effect van alle overige combinaties. Het effect van de televisie op maakwerk wordt even negatief ingeschat als het effect van geluidsmedia op leerwerk. De meest negatieve invloed heeft volgens de leerlingen het combineren van televisie met leerwerk.

Uit multiple regressieanalyse blijkt dat naast schooltype de denominatie van de school samenhangt met het gepercipieerde effect van een van de vier huiswerk-medium combinaties, namelijk de combinatie van televisie met maakwerk. Leerlingen van PC-scholen schatten het effect van televisie op maakwerk negatiever in dan leerlingen van RK- en openbare scholen ($\beta = 0,07$, $p < 0,01$).

Om meer inzicht te krijgen in de gepercipieerde effecten van de combinatie van huiswerk met televisie, zijn aan de leerlingen stellingen over de invloed van een zestal soorten televisieprogramma's op het maken van huiswerk voorgelegd. De respondenten moeten zich van elk type televisieprogramma voorstellen dat dit aanstaat terwijl zij bezig zijn met maakwerk of leerwerk en kiezen uit drie alternatieven: huiswerk gaat 'beter', 'slechter' of 'het maakt niet uit'. Een 6x2-variantieanalyse met twee binnen-proefpersonen-factoren, namelijk Programmatype (muziekprogramma's, nieuwsprogramma's, sportprogramma's, praatprogramma's, spel- of quiz-programma's, en dramaseries) en Huiswerktype (maakwerk en leerwerk) leidt tot hoofdeffecten van zowel programmatype als huiswerktype. Tabel 6 laat opnieuw zien dat leerlingen het effect van televisieprogramma's op leerwerk negatiever inschatten dan het effect van televisieprogramma's op maakwerk, $F(1,1671) = 794,86$, $p < 0,001$. Daarnaast is het gepercipieerde effect van televisie op maak- en leerwerk afhankelijk van het soort televisieprogramma, $F(5,8355) = 281,80$, $p < 0,001$. Blijkens een post-hoc-vergelijking met behulp van de Scheffé-test ($p < 0,05$) wordt het effect van muziekprogramma's op de televisie als significant minder negatief gezien dan het effect van de overige televisieprogrammatypen, namelijk nieuwsprogramma's, sportprogramma's, praatprogramma's, spel- en quiz-programma's en dramaseries. De onderlin-

Tabel 6. Gemiddelde gepercipieerde effecten van soorten televisieprogramma's op maakwerk en leerwerk.

Programmatype	maakwerk	leerwerk	**totaal**
muziekprogramma's	+0,16	–0,40	**–0,12**
nieuws	–0,33	–0,54	**–0,44**
sportprogramma's	–0,33	–0,57	**–0,45**
praatprogramma's	–0,39	–0,61	**–0,50**
spel/quiz	–0,46	–0,67	**–0,57**
dramaseries	–0,58	–0,75	**–0,67**
totaal	**–0,32**	**–0,59**	**–0,46**

Noot: De scores variëren van –1 (combinatie heeft sterk negatief effect) tot +1 (combinatie heeft sterk positief effect)

ge verschillen tussen de laatstgenoemde vijf programmatypen zijn in het algemeen niet significant, zij het dat het effect van dramaseries negatiever wordt ingeschat dat het effect van nieuws- en sportprogramma's. Er is echter ook een significante interactie tussen programmatype en huiswerktype, $F(5,8355) = 114{,}03$, $p < 0{,}001$. Dit interactie-effect houdt in dat de categorie muziekprogramma's in vergelijking met de andere programmatypen een bijzondere positie inneemt. Het effect van muziekprogramma's op leerwerk wordt aanzienlijk negatiever beoordeeld dan het effect van muziekprogramma's op maakwerk (het verschil tussen de gemiddelde scores is 0,56). Bij de andere programmatypen zijn de verschillen in gepercipieerde effecten tussen maakwerk en leerwerk weliswaar in dezelfde richting maar aanzienlijk kleiner (tussen 0,17 en 0,24).

DISCUSSIE

Geluidsmedia worden door middelbare scholieren veelvuldig als achtergrond bij het maken van huiswerk gebruikt, vooral bij het maakwerk. Tijdens de 5 à 6 uur per week die door deze groep gemiddeld aan leerwerk wordt besteed staan geluidsmedia 'soms' tot 'vaak' aan, terwijl maakwerk, waaraan wekelijks gemiddeld 5 uur wordt besteed, 'vaak' tot 'heel vaak' met geluidsmedia wordt gecombineerd. De betrokken media zijn meestal de cassette- of cd-speler en de radio, waarover bijna alle leerlingen zowel in de eigen kamer als in de huiskamer beschikken. De geluidsmedia worden tijdens het huiswerk als muziekmedia gebruikt, met als favoriete muzieksoorten pop, house en rock. De walkman wordt tijdens het huiswerk relatief weinig gebruikt, door minder dan 20% van de leerlingen, hoewel ook dit medium de meeste leerlingen ter beschikking staat. Misschien is het geluid via een hoofdtelefoon te indringend om huiswerk bij te maken.

In vergelijking met geluidsmedia wordt de televisie minder frequent als achtergrondmedium bij het huiswerk gebruikt. Gemiddeld genomen wordt zowel maakwerk als leerwerk 'soms' tot 'vaak' met de televisie gecombineerd, doch maakwerk vaker dan leerwerk. Indien het huiswerk met televisieprogramma's wordt gecombineerd gaat het meestal om muziekprogramma's en, in mindere mate, dramaseries.

Het gebruik van geluidsmedia en televisie tijdens het huiswerk correspondeert met de gepercipieerde effecten van de beide media op maak- en leerwerk: naarmate leerlingen aan een combinatie van huiswerk met media minder gunstige effecten op de huiswerkprestaties toeschrijven wordt die combinatie minder frequent gebruikt. Dit suggereert dat de leerlingen zich bij hun beslissing om al dan niet van achtergrondmedia gebruik te maken laten leiden door hun perceptie van het effect dat media op huiswerk hebben. Het is echter ook mogelijk dat de inschatting van het effect van achtergrondmedia wordt bepaald door het eigen huiswerkgedrag. Leerlingen die hun huiswerk veelvuldig met media combineren zijn wellicht geneigd het effect van media op huiswerk rooskleurig in te schatten.

Voor zowel maakwerk als leerwerk geldt dat de invloed van televisie op huiswerk negatiever wordt ingeschat dan de invloed van geluidsmedia. In experimenteel onderzoek is nooit een vergelijking gemaakt tussen het effect van achtergrondmuziek en dat van achtergrondtelevisie, maar het is aannemelijk dat televisie feitelijk hinderlijker is omdat dit medium niet alleen auditief afleidt maar de leerling tevens kan doen wegkijken van het huiswerk. De gedachte dat de televisie hinderlijker is dan geluidsmedia wordt ondersteund door de bevinding dat het wisselend aandacht besteden aan huiswerk en achtergrondmedium bij de televisie vaker voorkomt dan bij geluidsmedia. Bij geluidsmedia vormt het huiswerk voor de meerderheid van de respondenten de primaire taak, terwijl men de geluidsmedia op de achtergrond hoort. Bij de televisie komt het vaker voor dat leerlingen hun huiswerk onderbreken om hun aandacht even helemaal op het achtergrondmedium te richten.

De leerlingen menen dat maakwerk minder van achtergrondtelevisie te lijden heeft dan leerwerk. Of zij op dit punt gelijk hebben, is niet onderzocht. Wel is in de tot op heden verrichte experimentele studies gevonden dat complexe taken die continue concentratie vereisen meer

interferentie van achtergrondtelevisie ondervinden dan kortdurende eenvoudige problemen (Armstrong, 1993). Deze bevinding zegt echter weinig over het onderscheid tussen maakwerk en leerwerk, omdat beide huiswerktypen zowel complex als eenvoudig kunnen zijn. De perceptie van scholieren dat maakwerk beter met achtergrondtelevisie is te combineren dan leerwerk is vermoedelijk te verklaren doordat zij bij maakwerk vooral aan eenvoudig maakwerk denken en bij leerwerk vooral aan veeleisende leertaken.

De respondenten maken wat betreft de gepercipieerde effecten een onderscheid tussen verschillende typen televisieprogramma's. Het minst negatieve effect verwacht men van muziekprogramma's. In overeenstemming hiermee is dat muziekprogramma's de meest gebruikte televisieachtergrond zijn. Het meest negatieve effect gaat volgens de leerlingen uit van dramaseries, hoewel dit na muziekprogramma's de meest gekozen televisieachtergrond tijdens het huiswerk is. Kennelijk laten de leerlingen hun keuze van televisieprogramma's tijdens het huiswerk niet uitsluitend bepalen door de gepercipieerde effecten van die programma's. Dramaseries worden vermoedelijk aangezet omdat men de serie niet wil missen, waarbij de (gepercipieerde) negatieve invloed op de koop toe wordt genomen. Niet onderzocht is of er ook feitelijk een verschil in effect tussen muziekprogramma's en dramaseries bestaat, hoewel dit wel aannemelijk lijkt. Slechts in twee van de tot nu toe verrichte experimentele studies naar het effect van achtergrondtelevisie is de inhoud van de achtergrondtelevisie gevarieerd. In de ene studie (Armstrong & Greenberg, 1990) werden de effecten van soap en actiedrama vergeleken, in de andere (Armstrong et al., 1991) die van soap en reclame. In deze studies werden geen aanwijzingen gevonden dat de aard van de achtergrondtelevisie verschil maakt voor het effect op cognitieve taken.

Volgens de respondenten heeft de meest toegepaste combinatie – maakwerk gecombineerd met geluidsmedia – geen negatieve invloed op de huiswerkresultaten. Daarentegen menen de leerlingen dat de minder frequent gebruikte combinatie van leerwerk met geluidsmedia ongunstig op de huiswerkresultaten kan uitwerken. Het is echter onbekend of achtergrondgeluid inderdaad onschuldig is bij maakwerk en alleen op leerwerk negatief uitwerkt. Wel is experimenteel nagegaan wat het effect is van achtergrondgeluiden op met maak- en leerwerk verwante taken als begrijpend lezen (Martin, Wogalter & Forlano, 1988) en het memoriseren van getallen (Salamé & Baddeley, 1989). De in de experimenten gehanteerde opdracht tot begrijpend lezen is te zien als maakwerk, zij het dat de taak veel korter duurde dan een normale huiswerktaak. Het memoriseren van getallen valt in onze categorie leerwerk met de aantekening dat de overeenstemming met normaal huiswerk gering is. De experimentele opdracht tot begrijpend lezen blijkt uitsluitend gehinderd te worden door betekenisvolle achtergrondgeluiden, zoals gesproken taal en vocale muziek, terwijl instrumentale muziek niet van invloed is. Het memoriseren van getallen, een vorm van leerwerk, ondervindt daarentegen niet alleen hinder van tekst maar ook van instrumentale muziek. Generalisering van de experimentele taken naar leer- en maakwerk leidt tot de conclusie dat alle achtergrondmuziek een negatief effect kan hebben op leerwerk, maar dat maakwerk alleen gehinderd wordt door achtergrondmuziek met begrijpelijke teksten. Analoog aan de hierboven gevoerde discussie van de effecten van achtergrondtelevisie, valt het echter te betwijfelen of deze generalisering over het effect van achtergrondgeluiden op maak- en leerwerk stand zou houden indien de experimenten werden uitgevoerd met andere vormen van maak- en leerwerk, met name complex maakwerk en eenvoudig leerwerk.

Om na te gaan in hoeverre de televisie en geluidsmedia feitelijk effect hebben op huiswerkprestaties van (Nederlandse) scholieren is nieuw experimenteel onderzoek nodig. Het verrichte experimentele onderzoek naar de invloed van achtergrondgeluiden en -televisie suggereert dat daarin complexe taken die continue concentratie vereisen afgezet moeten worden tegen eenvoudige taken. Daarbij is van belang dat de experimentele taken, evenals dit bij huiswerk het geval is, enige tijd in beslag nemen. Op basis van het hier gepresenteerde enquête-onderzoek dient in het onderzoek naar de effecten van geluidsmedia vooral aandacht te worden besteed aan de effecten van pop-, house-, en rockmuziek, terwijl in onderzoek naar achtergrondtelevisie het effect van muziekprogramma's en dramaseries van belang is. Een aandachtspunt daarbij is in hoeverre de teksten van zowel dramaseries als de vocale achtergrondmuziek voor de proefper-

sonen begrijpelijk zijn. Onze verwachting is dat achtergrondstimuli die nauwelijks aandacht trekken, zoals overbekende muziek, zich zonder hinderlijke effecten laten combineren met het maken van huiswerktaken die eveneens weinig aandacht vereisen. Daarentegen zullen aandachtvereisende media-inhouden, zoals een interessante dramaserie op televisie, en inspannende huiswerktaken, zoals het lezen van een moeilijke tekst, wel met elkaar interfereren.

LITERATUUR

Armstrong, G.B. (1993). Cognitive interference from background television: Structural effects on verbal and spatial processing. *Communication Studies, 44*, 56-70.

Armstrong, G.B., Boiarski, G.A., & Mares, M.L. (1991). Background television and reading performance. *Communication Monographs, 58*, 235-253.

Armstrong, G.B., & Greenberg, B.S. (1990). Background television as an inhibitor of cognitive processing. *Human Communication Research, 16*, 355-386.

CBS (1993). *Statistiek van het voortgezet onderwijs 1991/'92: Scholen, leerlingen en examens*. Voorburg/ Heerlen: Centraal Bureau voor de Statistiek.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.

Martin, R.C., Wogalter, M.S., & Forlano, J.G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language, 27*, 382-398.

Neuwahl, N.M.E., & Van den Bogaart, P.H.M. (1989). *Huiswerk*. Leiden: Martinus Nijhoff.

Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology, 41A*, 123-140.

Manuscript ontvangen 26-10-1994
Geaccepteerd voor publikatie 14-12-1994

APPENDIX

Effectperceptieschaal

De gepercipieerde effecten van de vier huiswerk/medium-combinaties zijn vastgesteld met behulp van de volgende negen stellingen die voor elke combinatie zijn herhaald. Hieronder staan de antwoordalternatieven voor de combinatie televisie/maakwerk.

1. Televisie bij *maakwerk* zorgt ervoor
   O dat het maakwerk sneller af is
   O dat het maakwerk minder snel af is
   O maakt niet uit

2. O dat je tijdens het maakwerk vaker stopt
   O dat je tijdens het maakwerk minder vaak stopt
   O maakt niet uit

3. O dat je maakwerk beter doet
   O dat je maakwerk slechter doet
   O maakt niet uit

4. O dat je er meer van herinnert
   O dat je er minder van herinnert
   O maakt niet uit

5. O dat je je beter concentreert
   O dat je je slechter concentreert
   O maakt niet uit

6. O dat je betere cijfers haalt
   O dat je slechtere cijfers haalt
   O maakt niet uit

7. O dat je het beter begrijpt
   O dat je het slechter begrijpt
   O maakt niet uit

8. O dat je beter gaat nadenken
   O dat je slechter gaat nadenken
   O maakt niet uit

9. O dat je meer plezier hebt in huiswerk
   O dat je minder plezier hebt in huiswerk
   O maakt niet uit

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 2, pp. 191-193

Boekbespreking

Dolmans, D. (1994). *How students learn in a problem-based curriculum.* Proefschrift Rijksuniversiteit Limburg. Maastricht: Universitaire Pers Maastricht. ISBN: 90-5278-130-3 (f 41,-).

"De vanzelfsprekendheid waarmee conventionele curricula verondersteld worden het leren van studenten te bevorderen, lijkt omgekeerd evenredig aan de verantwoording die innovatieve curricula moeten afleggen over het leren van studenten in die programma's."

Waarschijnlijk bedoelt de auteur van deze stelling te zeggen, dat er een schril contrast bestaat tussen de verantwoording die van innovatieve curricula gevraagd wordt en de vanzelfsprekendheid waarmee conventionele curricula geaccepteerd worden. Deze stelling, behorend bij het proefschrift van Dolmans, typeert goed de achterdochtigheid waarmee nieuwe onderwijsconcepten tegemoet worden getreden. Het zou echter jammer zijn wanneer de tegenstelling waar de stelling op doelt zou worden opgelost door van innovatieve curricula minder verantwoording te verwachten. Juist het moeten aantonen dat iets nieuws ook, of misschien wel beter werkt dan iets waaraan men gewend is, is de drijfveer geweest achter het Maastrichtse onderzoek naar het leren van studenten in Probleem Gestuurd Onderwijs (PGO). Het proefschrift van Dolmans maakt deel uit van een onderzoekprogramma waarin diverse factoren die het leren van studenten in PGO beïnvloeden worden onderzocht, zoals de kwaliteit van de gebruikte problemen, de rol van tutoren, de rol van het activeren van voorkennis, enz. Het programma heeft reeds vele waardevolle inzichten en publikaties opgeleverd over het zelfstandig leren door studenten, en het proefschrift van Dolmans doet dat ook.

Want zelfstandig leren en PGO horen bij elkaar. Een belangrijk uitgangspunt bij PGO is dat onderwijs erop gericht moet zijn studenten te leren hoe ze in de toekomst met problemen om kunnen gaan en hoe ze hun verdere leven zelfstandig nieuwe kennis kunnen blijven verwerven. Eén van de kenmerken van PGO is dat de leerstof niet wordt voorgeschreven door de docent, maar dat studenten zelf moeten kiezen welke leerstof ze bestuderen om een probleem te kunnen verklaren. Een dilemma hierbij is dat de faculteit ook wil voorkomen dat er bij studenten lacunes in hun kennis ontstaan of blijven bestaan. Deze spanning tussen vrijheid bieden aan studenten om het zelfstandig leren te bevorderen enerzijds en garant staan voor de kennis die zij verwerven anderzijds is aanleiding geweest voor het onderzoek van Dolmans.

Dolmans heeft onderzoek gedaan naar het leren van medische studenten in PGO. Het Engelstalige proefschrift is opgebouwd rond vijf artikelen die in de internationale, vooral medische vakliteratuur zijn gepubliceerd of ter publikatie zijn aangeboden. Een algemene inleiding en een afsluitende discussie zijn eraan toegevoegd.

Vereenvoudigd weergegeven verloopt het model van PGO ongeveer als volgt. Studenten worden geconfronteerd met een serie problemen. Een probleem bestaat gewoonlijk uit een beschrijving van een verschijnsel waarvoor een verklaring gezocht moet worden. Wanneer docenten problemen ontwerpen, hebben ze bepaalde onderwijsdoelen voor ogen. Een probleem wordt vervolgens bediscussieerd in een onderwijsgroep met 8 tot 10 studenten onder begeleiding van een tutor. Gedurende de discussie komen vragen naar voren die opheldering behoeven. Deze vragen worden door de studenten omgezet in leerdoelen. De leerdoelen worden verondersteld de leeractiviteiten van studenten tijdens hun zelfstudie te sturen. Tijdens de zelfstudie zoeken studenten naar relevante informatie om de vragen die zijn vervat in de leerdoelen te kunnen beantwoorden. De studenten komen dan weer bij elkaar in de onderwijsgroep om de nieuw verworven kennis samen te brengen. Wat studenten hebben geleerd wordt uiteindelijk gemeten met een prestatietoets.

Dolmans heeft onderdelen van dit ideaaltypisch verloop van het leerproces in PGO aan de empirie getoetst. Ze heeft de relatie tussen verschillende elementen van het PGO-model onderzocht, met name die tussen onderwijsdoelen en de leerdoelen van studenten en die tussen leerdoelen en de leeractiviteiten die studenten tijdens hun zelfstudie ondernemen. Hierbij moet wel worden opgemerkt dat zowel 'leerdoelen' als 'leeractiviteiten' in het proefschrift worden

beperkt tot 'leerinhouden' (onderwerpen) die studenten kiezen en/of bestuderen. Het begrip 'leeractiviteit' heeft hier dus geen betrekking op bijvoorbeeld de aard van de mentale activiteiten die studenten verrichten om leerinhouden te verwerken, maar alleen op welke leerinhouden ze verwerken. Dolmans heeft gekeken naar *wàt* studenten bestuderen, niet naar *hoe* ze dat doen.

Wat betreft de relaties tussen onderwijs- en leerdoelen bleek dat ongeveer 15% van de onderwijsdoelen die docenten voor ogen stonden niet door de studenten werden geformuleerd als leerdoelen. Een classificatieprocedure wees uit dat dit vooral onderwijsdoelen waren die niet zoveel met de problemen te maken hadden maar meer met andere onderdelen van het curriculum, of die op meer dan één probleem betrekking hadden, of die te maken hadden met psychologische of sociologische onderwerpen. Bovendien bleek dat ongeveer 6% van de leerdoelen die studenten formuleerden niet overeenkwamen met enig onderwijsdoel. Deze leerdoelen hadden vooral te maken met deficiënties in voorkennis, toepassing van kennis in medische interventies (komt pas later in het curriculum aan de orde), en persoonlijke interesses in bepaalde onderwerpen. Al met al een redelijk hoge correspondentie tussen onderwijsdoelen en leerdoelen. Hierbij moet wel worden opgemerkt dat alle analyses op het niveau van de onderwijsgroep zijn gebeurd, en dus geen betrekking hebben op individuele studenten.

Om de relaties tussen leerdoelen en leeractiviteiten te kunnen onderzoeken heeft Dolmans de TOpic Checklist (TOC) ontworpen. Dit instrument bestond uit een inventarisatie van alle onderwerpen waarvan de docenten verwachtten dat studenten die in een bepaalde cursus bestudeerden. Studenten werd aan het eind van de cursus gevraagd hoeveel tijd zij aan elk onderwerp hadden besteed en in hoeverre ze het idee hadden elk onderwerp te beheersen. Per probleem werden deze twee maten vergeleken met de leerdoelen van de studenten. Ze bleken nauwelijks te correleren. Dolmans concludeert hieruit dat wat studenten van plan zijn te doen, zoals tot uitdrukking komt in de onderwerpen geformuleerd in de leerdoelen, geen relatie vertoont met de onderwerpen die ze in werkelijkheid bestuderen. Studenten doen kennelijk vaak niet wat ze zich voornemen en doen ook dingen die ze zich niet hebben voorgenomen. Voor deze verrassende uitkomst geeft de auteur tien mogelijke verklaringen.

De logische volgende vraag is dan natuurlijk: waardoor worden die leeractiviteiten van studenten dan wel gestuurd? In een interview- en vragenlijststudie probeert Dolmans hierop een antwoord te geven. Studenten van alle vier studiejaren werd een vragenlijst voorgelegd over de invloed van allerlei curriculumgebonden factoren op de zelfstudie-activiteiten van studenten. Factoren als de discussie in de onderwijsgroep, de leerdoelen, literatuurverwijzingen, blokdoelstellingen, lezingen, toetsvragen, andere studenten en tutoren bleken allemaal van invloed te zijn op de onderwerpen die studenten bestuderen. Haar analyses geven geen uitsluitsel over het relatieve gewicht van deze factoren bij de keuzeprocessen van studenten, of over de vraag of er groepen studenten zijn te onderscheiden die van elkaar verschillen in de wijze waarop ze deze bronnen gebruiken. De gehanteerde statistiek in dit onderzoek is ook wel erg eenvoudig: op alle 33 vragen van de vragenlijst werden afzonderlijke ANOVA's uitgevoerd met de vier studiejaren als onderscheidend criterium, zonder post-hoc analyses. Hoe bijvoorbeeld antwoorden op verschillende vragen met elkaar samenhangen en of bepaalde factoren een grotere invloed uitoefenen dan andere valt op deze manier dan ook niet te achterhalen.

Wel bleek bij eerstejaars-studenten de invloed van literatuurverwijzingen, lezingen en toetsen groter dan bij ouderejaars-studenten, terwijl de invloed van discussie in de onderwijsgroep en zelfgegenereerde leerdoelen groter was bij ouderejaars. Dolmans concludeert hieruit dat studenten in PGO meer zelfwerkzaam worden naarmate ze langer participeren in het curriculum. Maar er kan natuurlijk ook sprake zijn van selectieve uitval, waarbij vooral de niet-zelfwerkzame studenten in een vroegtijdig stadium hun studie staken. Alleen longitudinaal onderzoek kan hierover uitsluitsel geven.

Wat onderbelicht in het boek blijft het cognitief-psychologisch perspectief (Lowyck & Simons, 1991). Hiermee wordt bedoeld dat de leeractiviteiten die lerenden verrichten wellicht evenzeer of misschien wel meer gestuurd worden door factoren in de lerende (zoals voorkennis, interesse, persoonlijke doelen en motieven, voorkeuren voor leerstrategieën, opvattingen over leren en onderwijs, en dergelijke) dan door factoren van buiten (verwijzingen, doelstellingen,

toetsvragen, docenten, e.d.). In verschillende van de onderzoeken die Dolmans rapporteert komt bijvoorbeeld naar voren dat medische studenten een voorkeur hebben voor feitgerichte, biologische en fysiologische onderwerpen boven psychologische en sociologische onderwerpen die door hen als 'vaag' worden omschreven. Hierachter kan een wereld van opvattingen schuilgaan over de aard en waarde van kennis die het leergedrag van studenten sterk kan sturen. Op verschillende plaatsen komt ook de invloed van persoonlijke interesses op het kiezen en bestuderen van onderwerpen naar voren. Waar Dolmans in haar laatste onderzoek op zoek gaat naar andere variabelen die de keuze van leeractiviteiten bepalen dan de onderwijsdoelen, kijkt ze jammer genoeg alleen naar de curriculumgebonden factoren. Juist bij PGO, waarin zelfstandig leren zo'n centrale plaats inneemt, zou deze cognitief-psychologische bril heel interessant zijn om door te kijken.

Onderwijsdoelen zoals geformuleerd door de docenten zijn dus maar in beperkte mate van invloed op hetgeen studenten bestuderen. Eén van de door Dolmans geïnterviewde studenten verwoordt deze invloed als volgt: "The day before the test takes place, I read the course objectives to obtain information about what we were expected to study during this course".

REFERENTIE

Lowyck, J., & Simons, P.R.J. (1991). Zelfinstructie: een cognitief-psychologisch perspectief. Inleiding op het themanummer. *Pedagogische Studiën, 68,* 281-283.

J. Vermunt
Vakgroep Psychologie
Katholieke Universiteit Brabant,
en Instituut voor de Lerarenopleiding
Universiteit van Amsterdam

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 2, pp. 194-196

Boekbespreking

Geert ten Dam, Edith van Eck en Monique Volman.
Onderwijs en Sekse. Een verkenning van researchprogramma's. Trendrapport. 's Gravenhage: Directie Coördinatie Emancipatiebeleid, Stimuleringsgroep Emancipatie-Onderzoek, 1992 (te bestellen bij Distributiecentrum DOP, Postbus 11594, 2502 AN 's-Gravenhage, tel. 071-352500).

Een van de opmerkelijkste ontwikkelingen in de maatschappelijke ongelijkheid van onderwijskansen is dat vrouwen en meisjes hun onderwijsachterstand wat betreft eindniveau hebben ingehaald. Was het in de jaren zestig nog zo dat meisjes, ongeacht het sociaal milieu, lager scoorden in het bereikte onderwijsniveau dan jongens; in de jaren negentig is het verschil wat betreft dit eindniveau bijna geheel verdwenen. Een van de opmerkelijkste zaken in het onderzoek naar deze kansenongelijkheid is dat we over de oorzaken van deze afname (die al voor het begin van de tweede feministische golf een aanvang nam) nog grotendeels in het duister tasten. De onderzoeksinspanning op dit terrein is kennelijk een afgeleide van de hoofdinteresse van schoolloopbanen-onderzoekers, namelijk de milieu-ongelijkheid. Even opmerkelijk is overigens het empirische voortbestaan van horizontale verschillen in vakkenkeuze en de slechtere arbeidsmarktperspectieven. Gezien tegen deze achtergrond is de gedegen studie van Ten Dam en collega's een belangrijk werk omdat het niet alleen inzicht geeft in de weg die het sekse-ongelijkheidsonderzoek tussen 1982 en 1990 heeft afgelegd, maar ook aandacht besteedt aan het al dan niet in elkaar grijpen van theorie en empirie op dit gebied. Dit op elkaar betrekken van enerzijds theoretische proposities en concepten en anderzijds resultaten van empirisch onderzoek is een noodzakelijke voorwaarde voor kennisvermeerdering die gericht is op beantwoording van vragen als de bovenstaande.

Voorafgaand aan de weergave van het verrichte onderzoek worden in deze trendstudie de wetenschapsinterne en -externe context besproken waarin het onderzoek geplaatst moet worden. Daarbij komen de programmering en financiering van het onderzoek aan de orde, alsmede enige theoretische invalshoeken die voor de ontwikkeling van het onderzoeksterrein van belang zijn. Vervolgens worden de cijfers over de onderwijsdeelname gepresenteerd en de beleidsontwikkelingen ten aanzien van sekse-ongelijkheid en onderwijs besproken. Het geanalyseerde onderzoek wordt in vijf categorieën ondergebracht:

- het beschrijvend onderzoek en het evaluatie-onderzoek naar het door de overheid gestimuleerde emancipatiebeleid in het onderwijs;
- het literatuuronderzoek dat in de jaren zeventig op gang kwam naar het thema sekse-verschil;
- het loopbaanonderzoek dat zich onder andere richt op schoolloopbanen van allochtone meisjes, studieloopbanen van vrouwen in de volwasseneneducatie en in het hoger onderwijs alsmede de loopbanen van vrouwelijke onderwijsgevenden;
- het onderzoek naar sekseverschillen in schoolse vaardigheden en onderwijsprestaties dat toegespitst wordt op de exacte vakken en technische richtingen en de keuzen, prestaties en attituden van meisjes daarbinnen;
- het onderzoek naar het microniveau van het onderwijs waarbij centraal staan de inhoud van leermiddelen, leren en didactiek, interactieprocessen in de klas en de groeperingsvorm van leerlingen.

De auteurs zijn in hun trendreport grondig te werk gegaan. Zij geven een becommentarieerd overzicht van de ontwikkelingen in de uitgangspunten, de gehanteerde methoden en vraagstellingen op ieder van deze deelterreinen alsmede van de resultaten en opbrengsten. Daar komt bij dat dit overzicht bovendien nog helder en goed geschreven is en zich ook in dit opzicht positief

onderscheidt van veel andere gortdroge analyses. Dit betekent dat dit overzichtswerk ons inziens ook zeer geschikt is voor opleidingsdoeleinden.

Een van de uitgangsvragen van het rapport vormt de vraag naar de kennisgroei. Uit de studie komt naar voren dat de wetenschappelijke relevantie van veel beleid-georiënteerd onderzoek niet bijster groot is. Er wordt nauwelijks antwoord gegeven op theoretisch relevante vragen. Van researchprogramma's is veelal geen sprake. Er is slechts een geringe continuïteit in het onderzoek en er vindt nauwelijks terugkoppeling plaats van de empirie naar de theorie. Bovendien is het beleidsonderzoek nogal conjunctuurafhankelijk. De aard en omvang van het onderzoek varieert met de prioriteit die de 'gender'-kwestie heeft op de politieke agenda. De laatste jaren lijkt de aandacht van beleidswege weer enigszins af te nemen.

Een van de aardige aspecten van dit type studies dat verschillende soorten onderzoek met elkaar confronteert, is dat expliciet gemaakt kan worden dat ook in het 'theorieloze' beleidsgerichte onderzoek toch wel degelijk allerlei impliciete theoretische noties aanwezig zijn met de nodige implicaties.

Uit de retrospectieve analyse van het verrichte onderzoek komt naar voren dat de overheid aanvankelijk de sekse-thematiek definieerde als het probleem van de achterstandspositie van meisjes terwijl de (beleids)onderzoekers meer geneigd waren uit te gaan van een culturele socialisatie-optiek. Daarbij is sekse-ongelijkheid niet zozeer een kwestie van verschillen in prestaties, maar van verschillen in attitudes en belangstelling welke samenhangen met de culturele socialisatie. Het type vragen dat gesteld wordt door het beleid blijkt fors te verschillen van de vragen die bij vrouwenstudies naar voren komen. De conceptualisering van het probleem van de sekse-ongelijkheid in het onderwijs als een 'achterstands'- of arbeidsmarktprobleem leidt er bijvoorbeeld toe dat de onderzoeksaandacht sterk gericht wordt op deelname aan exacte vakken en technische richtingen, waarbij de culturele aspecten van dit thema niet altijd de aandacht krijgen die zij verdienen. Tussen haakjes kan worden geconstateerd dat ook op beleidsniveau voorstellen voor een inhoudelijke overbrugging van de 'two cultures' (C.P.Snow) vrijwel ontbreken (vergelijk "Kies Exact').

Bij de analyse en vergelijking van de verichte research zijn de opstellers van de studie uitgegaan van de zogenaamde 'drie-lagen-methodologie' van Hetebrij (1983). De daarmee samenhangende methode van reconstructie van researchprogramma's is erop gericht wegen aan te geven waarlangs een wetenschapsgebied verder ontwikkeld kan worden. Naast de twee lagen van het theoretisch-conceptuele onderzoek en het empirisch onderzoek wordt hierbij een tussenliggende laag onderscheiden die de beide eerste lagen verbindt door een geëxpliciteerde theoretische oriëntatie: een researchprogramma. In de trendstudie blijkt deze lagen-benadering goed te functioneren als beschrijvings- en ordeningskader voor de vraag naar de kennisvermeerdering. De resultaten overziende is het totaalbeeld niet al te rooskleurig. Men treft slechts enkele aanzetten tot researchprogramma's aan. Meestal is er sprake van een gescheiden opereren van theoretisch en empirisch onderzoek. De meer specifieke vragen waartoe het drie-lagen-schema aanleiding geeft (bijvoorbeeld de relatie tussen de theoretische oriëntatie en de keuze voor bepaalde onderzoeksmethoden) konden in deze reconstructie van het uitgevoerde onderzoek nauwelijks worden beantwoord. Dit neemt niet weg dat de overzichtelijke studie veel aanleiding geeft om na te denken over hoe het verder moet met dit onderzoek en welke onderzoeksaanzetten (opnieuw) onze aandacht behoeven. Uit de trendstudie komt duidelijk naar voren dat bepaalde kennisidealen in de verste verten niet benaderd worden. Denk bijvoorbeeld aan het kennisideaal dat zich toespitst op het ontwikkelen van theorieën en vervolgens op het concreet onderzoeken van de schoolinterne mechanismen en socialisatieprocessen die maken dat meisjes anders of ongunstig kiezen en/of presteren. Onderzoek richt zich voornamelijk op het vaststellen van factoren die samenhangen met bepaalde keuzen en prestaties, zonder dat theorieën ontwikkeld worden die deze samenhang verklaren. Anderzijds wordt duidelijk dat het voor de versterking en verbetering van dit onderzoek nuttig is expliciet aan te sluiten bij bestaande onderzoekstradities zoals het schooloopbanenonderzoek en het effectieve school-onderzoek. De trendstudie bevat een aantal interessante aanbevelingen voor toekomstig onderzoek, waarbij ook ge-

1015912

bruik gemaakt is van de mening van een aantal deskundigen Kortom een opmerkelijke studie die vanuit de belangrijkste vraag, namelijk die van de kennisgroei, niet alleen achteruit maar ook vooruit kijkt. Het ware wenselijk dat zulke heldere overzichten ook op vele andere onderwijsprobleemgebieden tot stand zouden komen.

LITERATUUR

Hetebrij, M. (1983). *De ontwikkeling van researchprogramma's als methodisch probleem.* Amsterdam: VU Uitgeverij.

Cees A. Klaassen
Vakgroep Onderwijskunde
K.U. Nijmegen

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 3, pp. 197-199

Leren en onderwijzen van strategieën in het talenonderwijs

Gert Rijlaarsdam
Instituut voor de Lerarenopleiding, Universiteit van Amsterdam

BIJ WIJZE VAN INLEIDING

Het onderwijs in Nederland is in beweging en het talenonderwijs blijft daar niet bij achter. De basisvorming met eindtermen en later kerndoelen kristalliseerde het vakdidactisch denken van één à twee decennia, bracht eenheid in de vakinhoud en instigeert, zij het minder expliciet, vernieuwingen in de didactiek. De basisvorming vereiste het opnieuw definiëren van de inhoud langs twee wegen: de inhoud moest intrinsiek relevant zijn ('de groeikern' van het vak) en extrinsiek (relevantie voor school, beroep en maatschappij). De doordenking van die twee eisen aan de inhoud heeft door de invoering van de basisvorming een nieuwe impuls gekregen en is zeker nog niet afgerond. Zowel voor het schoolvak Nederlands (zie bijvoorbeeld Rijlaarsdam 1989, Bonset et al. 1993, Sijtstra 1991) als voor de moderne vreemde talen (ook Sijtstra 1991) zijn pogingen ondernomen om de inhoud zo zuinig mogelijk te verkavelen. Maar een bevredigende indeling is nog niet voorhanden. Het onderwijs vraagt daar wel om. Meer leerlingen moeten meer talen leren in de basisvorming dan voorheen, in minder onderwijstijd: afstemming van inhoud en didactiek met het oog op positieve transfer wordt door ieder gewenst.

De nadruk in het nieuwe talenonderwijs ligt op het verwerven van een beperkt aantal taaltaken, ten behoeve van welomschreven communicatieve situaties. Het verwerven van de benodigde vaardigheden is zelf ook object van onderwijs aan het worden: het 'leren leren'-paradigma vindt ook in het talenonderwijs ingang. In de publikaties over Kerndoelen worden zowel voor Nederlands als voor de Moderne Vreemde Talen daarover expliciet opmerkingen gemaakt; bij Moderne Vreemde Talen wordt daar zelfs een kerndoel aan gewijd.

Dit vereist een dubbele bodem in de didactiek. Leerlingen moeten bijvoorbeeld niet alleen leesvaardig worden, via een procesgerichte didactiek, maar moeten ook *leren* leesvaardig te worden, dus leren in nieuwe lees- en leersituaties een beroep te doen op aanwezige strategieën om die wellicht op een nieuwe wijze in te zetten. Zij moeten, met ander woorden, niet alleen leren hun leesproces , maar ook hun lees-leer-proces te coördineren. Aan de docent en aan het leerboek stelt dit soort onderwijs heel nieuwe eisen.

Het 'leren-leren'-paradigma zien we ook in de vernieuwingsvoorstellen en de onderwijsprojecten in de tweede fase van het voortgezet onderwijs opgeld doen. De voorstellen voor de vernieuwing van het examenprogramma Nederlands (Braet & Hendrix 1991) en van het examenprogramma Moderne Vreemde Talen (Commissie-Van Els) leiden in beide gevallen tot herdefinities van het schoolvak en tot een sterke nadruk op vaardigheidsonderwijs, waarin de lerende ook moet leren de communicatieprocessen te controleren door bijvoorbeeld niet alleen in de rol van communicator te treden, maar ook in de rol van observator en evaluator (beide zijn leerdersrollen). Veel onderwijsprojecten in Nederland gaan over de aansluitingsproblematiek HAVO-HBO en VWO-WO: de kern van het overleg tussen docenten van de verschillende instellingen wordt gevormd door zowel de leerstof alsook door domeinspecifieke *leer*vaardigheden.

Een en ander leidt tot een zekere mate van ontvankelijkheid in het onderwijs voor psychologische invalshoeken. Men zoekt naar de uitvoerende en regulerende processen die ten grondslag liggen aan een goede taakuitvoering, en parallel daaraan naar de verwervingsprocessen (*leert-*

Adres: Herengracht 256, 1016 BV Amsterdam.

aakuitvoering en -regulering) (zie bijvoorbeeld Bonset et al., 1992, p.23). Ook in de taal- en literatuurwetenschap is deze invalshoek te bespeuren: niet langer is uitsluitend het systeem van de taal of de tekst object van onderzoek, maar ook de cognitieve processen van interpretatie en produktie van teksten. Daardoor komt (een deel van) de taal- en literatuurwetenschap veel dichter te staan bij onderwijspraktijk en onderwijsonderzoek. Dit themanummer geeft een kleine vlootschouw van inspanningen van Nederlands toegepast taalkundig en toegepast letterkundig onderzoek, varend onder de vlag van cognitieve strategieën in het taalonderwijs. Vragen die aan de orde komen zijn: (1) welke strategieën moeten leerlingen leren en welke worden onderwezen; (2) in hoeverre zijn de aangetoonde of onderwezen strategieën modusbepaald (lezen, schrijven, spreken, luisteren) of codespecifiek (geldig voor Nederlands als moedertaal, Nederlands als Tweede Taal of voor een Moderne Vreemde Taal); (3) hoe worden die strategieën onderwezen en hoe zouden die strategieën onderwezen moeten worden?

Van der Pool en Van Wijk presenteren een psycholinguïstisch model dat de processen beschrijft die ten grondslag liggen aan lezen en schrijven. Zij leggen in hun bijdrage de nadruk op de conceptuele componenten en op de overeenkomsten tussen lezen en schrijven. Vervolgens komt het begrip strategie aan de orde: 'de globale representatie van de procedure(s) om een bepaald doel te bereiken'. Vanuit de ontwikkelingspsychologie brengen zij het onderscheid aan tussen twee wijzen waarop taaltaken kunnen worden aangepakt: procedureel en strategisch handelen. Alleen de laatste wijze biedt de taakuitvoerder gelegenheid te leren van de taakuitvoering. Het taalonderwijs zou leerlingen dus moeten leren strategisch te handelen in plaats van louter procedureel. Van der Pool en Van Wijk besluiten hun artikel met de waarschuwing dat het gevaar groot is dat strategie-onderwijs verwordt tot procedure-onderwijs, en 'dan bereikt men bij de leerling wel een verrijking van de basale (spontane) vorm van informatieverwerking, maar benut men niet de mogelijkheid om meer complexe (aangeleerde) vormen van informatieverwerking te ontwikkelen'.

De bijdrage van Rijlaarsdam, Oostdam en Bimmel beschrijft een stand van zaken met betrekking tot strategie-onderwijs voor het basisonderwijs (Nederlands en Engels) en voor de eerste fase van het voortgezet onderwijs (Nederlands, Engels en Frans). In het door hen geponeerde model van taaltaakuitvoering en taal-leertaakuitvoering maken zij expliciet onderscheid tussen twee typen taken en processen: taaltaken en taal*leer*taken. Zij analyseerden schoolboeken op inhoudelijke en didactische aspecten: worden er taalgebruiksstrategieën en/of taalleerstrategieën aangeboden, en in hoeverre bevat de oefening elementen die zelfregulatie en transfer naar andere taken kan bevorderen? Voor het vak Nederlands zijn de resultaten nog enigszins bemoedigend, maar het onderwijs in Moderne Vreemde Talen vertoont (nog steeds) veel kenmerken van 'blind training'.

In het kader van een studie naar mogelijkheden om transferbevorderende onderwijsmaatregelen te treffen, rapporteren Schouten-van Parreren, De Glopper en Van Daalen-Kapteijns correlationeel onderzoek met betrekking tot woordleervaardigheden. De idee is dat wat leerlingen leren ten behoeve van de taalverwerving van taal A ook een rol zou moeten spelen bij het onderwijs in taal B. De auteurs laten zien dat de woordkennis in het Nederlands, Engels en Frans matig tot sterk samenhangt, dat woordkennis sterk samenhangt met de woordleervaardigheden 'afleiden van de betekenis uit de context' en 'analyseren van de morfologische structuur' en dat deze samenhang toeneemt met de leeftijd. Het heeft dus alle zin, stellen de auteurs, leerlingen bij Nederlands woordleervaardigheden te onderwijzen, onder condities die de waarschijnlijkheid maximaliseren dat die vaardigheden ook in andere, vreemde-taalsituaties, gebruikt zullen worden. 'Meta-woordkennis' kan bij dergelijk onderwijs echter een belangrijk belemmerende factor zijn. Want het blijkt zo te zijn dat een klein aantal leerlingen consistent in alle talen de eigen kennis ruimschoots overschat. Zij nemen veel woorden dus niet waar als onbekend, en ontzeggen zich zo de kans een nieuw woord te verwerven. Naast woordleervaardigheden zal het onderwijs dus ook aandacht moeten besteden aan meta-cognitieve aspecten van woorden leren.

Een eerste aanzet voor inzicht in effectieve strategieën in het schrijfonderwijs in de Moderne Vreemde Talen leveren Broekkamp en Van den Bergh. Een ervaringsfeit van docenten is dat leerlingen die in de vreemde taal schrijven zo gefixeerd raken op de taalvorm, dat zij aan andere,

globale tekststrategieën, niet meer toekomen. Broekkamp en Van den Bergh onderzochten via een experiment in hoeverre aandacht van schrijvers in een vreemde-taalsituatie beïnvloed wordt door kenmerken van de te reviseren tekst (aantal en type problemen) en revisiedoelen (impliciet of expliciet). Uit hun onderzoek blijkt dat proefpersonen zich sterk laten beïnvloeden door het aantal lagere-orde-problemen dat de tekst bevat en dat zij in dat geval minder hogere-orde-problemen detecteren en reviseren. Als het revisiedoel echter expliciet aandacht besteedt aan hogere-orde-aspecten van de tekst, besteden proefpersonen minder aandacht aan feilen in de tekst van lagere orde, zonder aantoonbaar effect echter op het aantal hogere-orde-problemen dat zij aanpakken. Broekkamp en Van den Bergh betogen dat in het vreemde-talenonderwijs leerlingen moeten leren hun aandacht over verschillende aspecten van de tekst te verdelen en hun preoccupatie met linguïstische feilen in te dammen.

Hacquebord kwam in haar onderzoek tot de vaststelling dat Turkse leerlingen in vergelijking met Nederlandse leerlingen goed scoren op tekstbegriptoetsen, maar laag op taalvaardigheidstests. Aan de hand van andere testgegevens maakt Hacquebord aannemelijk dat Turkse kinderen anders lezen dan Nederlandse kinderen. De Turkse leerlingen lezen waarschijnlijk meer 'top-down' dan Nederlandse leerlingen, die meer 'bottom-up' lezen. In een vervolgonderzoek onderwierp Hacquebord zwakke lezers aan een experimenteel leesprogramma, dat o.a. beoogde leerlingen te leren leesstrategieën toe te passen op verschillende tekstuele niveaus en tot zelfcontrolerend leesgedrag te brengen. Het leesprogramma bleek in korte tijd significante vooruitgang te bewerkstelligen bij zowel anderstalige als Nederlandstalige zwakke lezers, ook op enkele relevante deelvaardigheden. Waar het gaat om het aanbrengen van taalspecifieke vaardigheden schiet het programma in de ogen van de auteur nog te kort.

Andringa besluit dit themanummer met een bijdrage vanuit ontwikkelingspsychologisch gezichtspunt over interpretatiestrategieën bij het lezen van literaire teksten. Met behulp van hardopdenkprotollen toont zij aan dat gevorderde lezers op andere wijze verscheidene tekstlagen met elkaar in verband brengen dan minder gevorderde lezers. Andringa stelt vast dat deze groepen lezers verschillen in het vermogen simultaan voorstellingen op te bouwen en die op elkaar te betrekken en toont enkele cognitieve strategieën die hiermee gepaard gaan als het maken van inferenties en generalisaties. Dit onderzoeksterrein is zeer pril: de leespsychologie van literaire teksten is nog weinig ontwikkeld. Toch is het voor het literatuuronderwijs van groot belang dat inzicht in informatieverwerkingsprocessen verkregen wordt, teneinde leerlingen meer te kunnen bieden aan onderwijs dan de literaire teksten zelf, verwerkingsopdrachten (het befaamde 'boekverslag' volgens per docent wisselende inhoudslijstjes) en literatuurgeschiedenissen (zie Janssen 1992).

Ik hoop dat dit themanummer kan bijdragen tot een grote inspanning van het taalonderwijsonderzoek ten behoeve van de onderwijspraktijk, die ons allen zo ter harte gaat.

LITERATUUR

Braet, A. & Hendrix, T. (red.) (1991). *Het CVEN-Rapport. Eindverslag van de Commissie Vernieuwing Eindexamenprogramma's Nederlands taal en letterkunde v.w.o. en h.a.v.o.* Den Haag: Sdu.

Bonset, H., de Boer, M., & Ekens, T. (1992). *Nederlands in de basisvorming: een praktische didactiek.* Muiderberg: Coutinho.

Janssen, T. (1992). *Het literatuuronderwijs Nederlands in de bovenbouw van het HAVO en VWO.* Resultaten van een nationale enquête. Amsterdam: Instituut voor de Lerarenopleiding.

Rijlaarsdam, G.C.W. (1989). *Basisvorming Nederlands.* Leiden: Martinus Nijhoff.

Sijtstra, J.M. (1991). *Doel en inhoud van het taalonderwijs. De ontwikkeling van een model voor domeinbeschrijvingen van taalonderwijs.* (Diss. KUB). Arnhem: Cito.

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 3, pp. 200-214

Proces en strategie in een psycholinguïstisch model van schrijven en lezen

Els van der Pool en Carel van Wijk
Werkverband Tekstwetenschap, Faculteit der Letteren, K.U. Brabant

ABSTRACT

The conceptual, linguistic, and physical processes involved in writing and reading are discussed in terms of a psycholinguistic model that fits in with theoretical proposals made among others by Levelt (1989) and Van Dijk and Kintsch (1983). Special attention is paid to the role of strategies in these processes. For the sake of clarity, an analysis of the concept 'strategy' has been included. This analysis underscores the necessity to distinguish sharply between goals, strategies, and procedures (viz. the why, what, and how). In line with Bereiter and Scardamalia's (1987) model, developmental changes in writing and reading are characterized as a transition from procedural to strategic modes of information processing. The paper concludes with a number of considerations that need specific attention when designing a process-oriented language curriculum.

1. INLEIDING

Voor een gerichte, effectieve ontwikkeling van vaardigheden is het zinvol aandacht te besteden aan onderliggende mentale processen en strategieën (Van den Bergh & Meuffels, 1993). Dat daarvoor binnen het taalonderwijs en de taalbeheersing een toenemende belangstelling bestaat, blijkt onder andere uit de kerndoelen die geformuleerd zijn voor het voortgezet onderwijs (Commissie Herziening Eindtermen, 1990), en de groeiende aandacht van onderzoekers voor onderliggende cognitieve aspecten (Van den Bergh, Rijlaarsdam & Breetvelt, 1992).

In deze bijdrage wordt een psycholinguïstisch kader uitgewerkt ten behoeve van de theoretische onderbouwing van het proces-gericht taalonderwijs. Dergelijk onderwijs veronderstelt inzicht in de processen en strategieën die ten grondslag liggen aan de te leren vaardigheden. Aan de hand van een model van schrijven en lezen wordt in paragraaf 2 een karakterisering gegeven van de processen bij produktie en verwerking. Dit model vormt in zijn geheel de basis voor de bespreking van strategieën en vaardigheden. In paragraaf 3 staan strategieën centraal, in paragraaf 4 de ontwikkeling van vaardigheden en de relatie met een strategische werkwijze. Paragraaf 5 geeft aandachtspunten voor proces-gericht onderwijs waarin het accent ligt op strategisch werken.

2. EEN COGNITIEF MODEL VOOR DE PRODUKTIE EN VERWERKING VAN TEKST

Figuur 1 geeft schematisch weer welke procescomponenten en kennisbestanden een rol spelen bij de produktie en verwerking van tekst. Het model is gebaseerd op algemene modellen van de taalgebruiker en op meer specifieke modellen van produktie of verwerking van geschreven tekst.[1]

2.1 Globale karakterisering van schrijven en lezen

In het model worden zowel produktie als verwerking van tekst behandeld. Beide processen hebben een aantal aspecten gemeenschappelijk: zij maken bijvoorbeeld op grammatisch en fonologisch niveau gebruik van dezelfde kennisbestanden. Bovendien zijn ze operationeel nauw verweven: iedere schrijver is tegelijkertijd lezer bij het reviseren van zijn eigen tekst.

Adres: Postbus 90153, 5000 LE Tilburg.

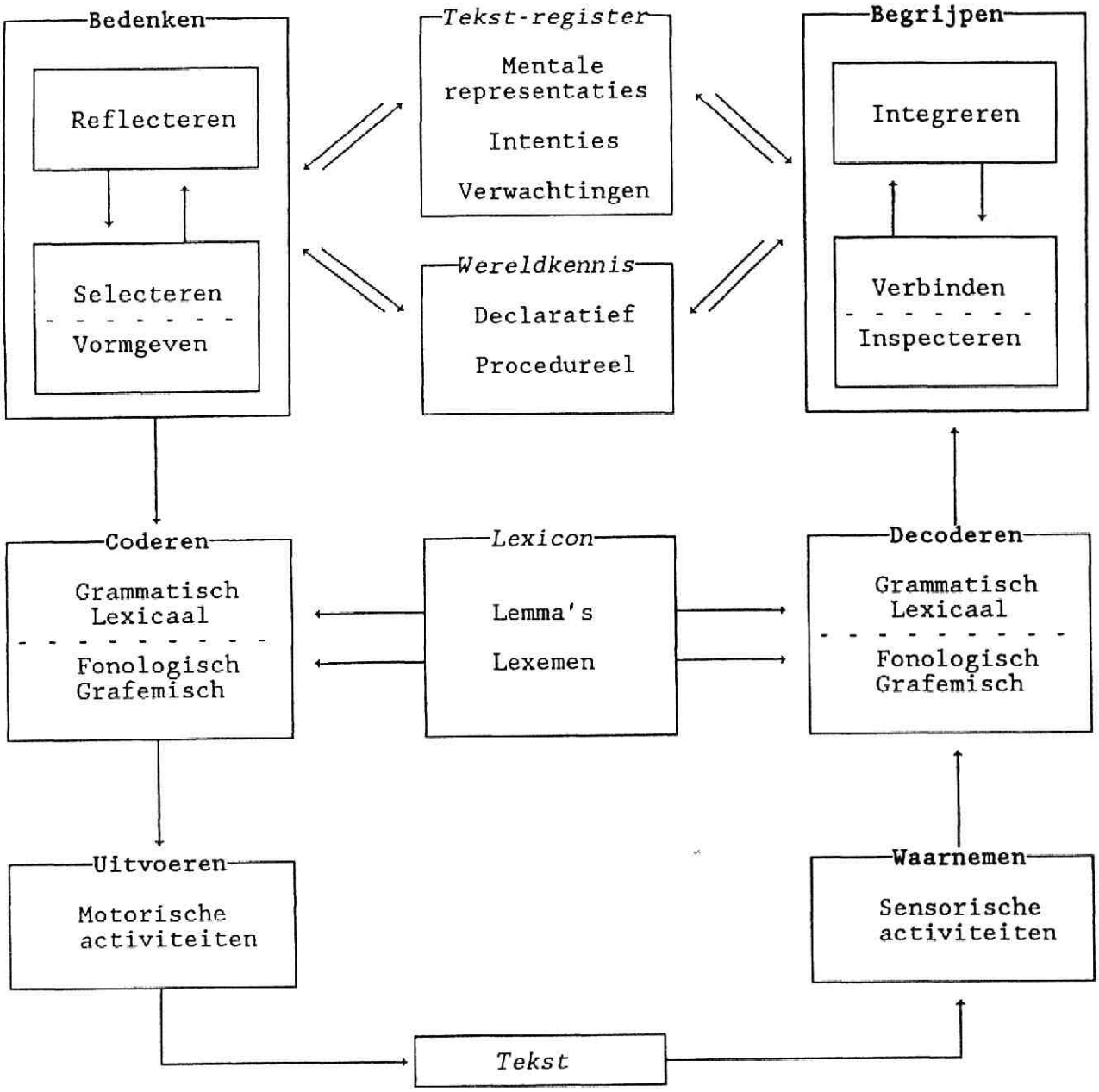


Fig. 1. Een cognitief model voor de produktie en verwerking van tekst.

Het schrijven en lezen van een tekst verloopt gefaseerd: er vinden achtereenvolgens diverse processen plaats in verschillende componenten.
Aan de linkerkant van figuur 1 staan van boven naar beneden de produktiecomponenten. Bij het schrijven wordt een boodschap bedacht, vervolgens talig gecodeerd en daarna door het uitvoeren van motorische activiteiten op papier gezet. Het resultaat is een tekst.

Aan de rechterkant van figuur 1 staan van beneden naar boven de verwerkingscomponenten. Tijdens het lezen worden door middel van sensorische activiteiten van het oog tekenreeksen waargenomen. De decodeerprocessen reconstrueren hieruit de lexicale invulling en syntactische vorm van eenheden die min of meer samenvallen met deelzinnen. In het begrijp-proces wordt iedere eenheid verbonden aan andere (reeds gelezen) eenheden en verrijkt met wereldkennis. Het resultaat is een mentale representatie.

De produktie- en verwerkingscomponenten zijn van verschillende typen: bedenken en be-

grijpen zijn *conceptueel*, coderen en decoderen zijn *talig*, en uitvoeren en waarnemen zijn *fysiek* (Dijkstra & Kempen, 1993).

Tekstregister
De intenties en verwachtingen die in een gegeven situatie het lees- en schrijfproces sturen, worden tijdelijk opgeslagen in een tekstregister (cf. 'discourse record' van Levelt, 1989, p. 111). Dit register bevat ook de mentale representaties die een lezer en een schrijver van een tekst maken.

Zowel een lezer als een schrijver creëren een mentale representatie van de tekst. De lezer bouwt een mentaal 'plaatje' op van de reeds verwerkte tekst, de schrijver van de geproduceerde c.q. nog te produceren tekst.

De intenties en verwachtingen vervullen een sturende rol in het lees- en schrijfproces. Bij het lezen bepaalt de intentie welke informatie een lezer onthoudt. Heeft een lezer bij een beschrijving van een huis de intentie het te kopen, dan zal hij andere informatie onthouden dan wanneer hij wil gaan inbreken (Anderson & Pichert, 1978). Bij het schrijven bepaalt de intentie inhoud en organisatie van de tekst. Zo zal een schrijver die de intentie heeft te amuseren, andere informatie selecteren dan wanneer zijn doel is een lezer te overtuigen (Matsuhashi, 1981).

Ook verwachtingen zijn van invloed op het lezen en schrijven. Bij het lezen zijn verwachtingen grotendeels gebaseerd op kennis en opgedane ervaring en hebben ze de vorm van stereotiepe patronen zoals scripts (Schank & Abelson, 1977). Ze worden onder andere geactiveerd door specifieke zinsneden. Zo verwacht een lezer dat na een frase als 'er was eens' een fantasievertelling volgt. Vervolgens roept zo'n frase verwachtingen op over de structuur en inhoud van de vertelling: hoofdpersonen worden geïntroduceerd (mensen, dieren of niet-bestaande wezens), 'goed en kwaad' raken met elkaar in conflict, maar gelukkig komt de redding net op tijd en dus 'leefden ze nog lang en gelukkig' (Propp, 1984).

Ook een schrijver heeft verwachtingen, bijvoorbeeld over de voorkennis, de achtergrond en het leesdoel van de potentiële lezer. Deze verwachtingen werken met name sturend bij het selecteren en vormgeven van informatie.

Wereldkennis
De conceptuele componenten maken gebruik van wereldkennis. Deze bestaat uit declaratieve kennis en procedurele kennis. De declaratieve kennis is te vergelijken met encyclopedische kennis die de vorm heeft van 'de burgemeester van Breda heet Nijpels' of 'Stefan Pettersson speelde bij Ajax'. Deze declaratieve kennis heeft betrekking op algemene feiten (propositionele kennis) en contexten (situationele kennis; Levelt, 1989, p. 10, 72). De procedurele kennis die een taalgebruiker ter beschikking staat, heeft de vorm van als-dan regels die worden toegepast op declaratieve kennis. Bijvoorbeeld 'Als een fractievoorzitter naar de achtergrond moet verdwijnen, maak hem dan burgemeester' of 'Als je een goede voetballer zoekt, neem dan een Ajaxied'.

Lexicon
De talige componenten doen een beroep op het lexicon. Voor ieder element in het lexicon is verschillende informatie aanwezig (Dijkstra & Kempen, 1993). In het 'lemma'-deel is syntactische en conceptuele informatie opgeslagen. Bijvoorbeeld 'geven' is een transitief werkwoord met drie casusrollen (agens, object, indirect object) en het kan de betekenis van 'aanreiken', 'toedienen' of 'verstrekken' hebben. In het 'lexeem'-deel wordt fonologische en morfologische informatie bewaard. Bijvoorbeeld bij 'geven' ligt de klemtoon op de eerste lettergreep; het is een sterk werkwoord, de vervoeging van de verleden tijd kent klinkerwisseling ('gaven').

Gebruikmakend van wereldkennis vinden er binnen de beide conceptuele componenten 'bedenken' en 'begrijpen' redeneerprocessen plaats die de taalgebruiker tot nieuwe inzichten doen komen (zie ook par. 2.3 en 4). Hierdoor is er geen eenrichtingverkeer waarbij kennis alleen

gebruikt wordt, zoals bij het lexicon het geval is. De conceptuele componenten kunnen op basis van redeneerprocessen veranderingen aanbrengen in de bestaande kennis. Vandaar dat er in het model niet alleen een pijl staat van kennis naar de conceptuele componenten die het gebruik symboliseert, maar eveneens een pijl naar wereldkennis toe die voor deze verandering staat.[3]

2.2 Karakterisering van de conceptuele componenten

Psycholinguïstisch onderzoek heeft zich met name gericht op de fysieke en talige componenten. Onderzoek naar coderen en uitvoeren en naar waarnemen en decoderen heeft een schat aan gegevens opgeleverd op basis waarvan deze onderdelen nader gespecificeerd zijn. De invulling van de conceptuele componenten is echter veel minder uitgekristalliseerd. Het onderzoek naar conceptuele processen is pas sinds de tachtiger jaren goed op gang gekomen (o.a. Just & Carpenter, 1987; Rayner & Pollatsek, 1989; Bereiter & Scardamalia, 1987; Flower & Hayes, 1981). Met de nu volgende karakterisering van de conceptuele componenten beogen we de theoretische voorstellen die momenteel opgang maken in samenhang en op concrete wijze te bespreken.

Bedenken van informatie

Tijdens het schrijven worden in de conceptuele component de intenties nader uitgewerkt (REFLECTEREN), de inhoud en organisatie van de informatie bedacht (SELECTEREN) en een presentatiestijl bepaald (VORMGEVEN). Deze processen zijn recursief en iteratief, ze wisselen elkaar af en kunnen worden herhaald (Hayes & Flower, 1980).

Stel dat iemand een sollicitatiebrief gaat schrijven met de intentie 'ik wil in aanmerking komen voor een gesprek'. Hoe verloopt dan het bedenken van informatie? Ter illustratie staat in tabel 1 een uitgewerkt voorbeeld voor één onderdeel van de brief: de argumentatie om voor de functie in aanmerking te komen.

Tijdens het **reflecteren** werkt de schrijver de intentie nader uit tot een of meer schrijfdoelen. Daarbij spelen naast verwachtingen ook een aantal aspecten van de taalgebruikssituatie een rol: het teksttype (het betreft een brief en geen gesprek), het onderwerp (het gaat om tijdelijk werk en geen vaste baan), de lezer (een commissielid en geen taaldocent) etcetera. Door te reflecteren op deze aspecten worden schrijfdoelen vastgesteld die functioneren als richtlijnen voor het selecteren en vormgeven. In het sollicitatievoorbeeld zijn mogelijke schrijfdoelen bijvoorbeeld een zelfbewuste presentatie geven, kundigheid laten blijken en, zij het in bescheiden mate, vooral positieve kanten benadrukken (Ia, Ib, Ic).

Deze schrijfdoelen sturen het selectie- en vormgevingsproces voor zover het voorafgaat aan het schrijven (de planning). Ze functioneren tevens als toetssteen bij het evalueren van de tekst (de revisie).[4]

Het **selecteren** omvat het ophalen en organiseren van informatie. Door opgehaalde informatie te relateren aan de schrijfdoelen wordt de geschiktheid om opgenomen te worden getest. Terwijl de inhoud wordt bepaald, worden ook keuzes gemaakt ten aanzien van de volgorde van de informatie. Dat bepaalt de structuur van de boodschap.

Zo moet een sollicitant afwegen welke informatie wel of juist niet geschikt is om op te nemen, en welke volgorde hij het meest doelmatig acht. De sollicitant kan besluiten bepaalde argumenten niet op te nemen in de brief. Gezien het feit dat de persoon zich graag van zijn beste kant wil laten zien, is het niet zo verstandig minder positieve punten te noemen, zoals zijn leeftijd (IId). Ten aanzien van zijn ervaring (IIc) kan de sollicitant besluiten om de relatieve eenzijdigheid ervan niet te vermelden.

Voor de informatie die hij wel opneemt, moet hij een geschikte volgorde kiezen. De sollicitant kan zijn bewering dat hij een geschikte kandidaat is onderbouwen, zoals (IIIa). Hij kan ook eerst de argumenten geven en vervolgens concluderen dat hij zeer geschikt is (IIIb). Omdat een van de schrijfdoelen is 'zelfbewust over komen', gaat in dit geval de voorkeur uit naar de meest directe aanpak, de onderbouwing van de geschiktheid (IIIa).

Tabel 1. Het conceptuele proces bij het schrijven van een argumentatief deel in een sollicitatiebrief.

Wat wil ik bereiken?
Intentie: Ik wil voor een gesprek uitgenodigd worden
Verwachting: De lezer zoekt punten die opvallen in de 'grijze massa'

Hoe kan ik dat met mijn brief realiseren?
Reflecteren: Ga mogelijke schrijfdoelen na
- Ia Geef zelfbewuste presentatie
- Ib Laat kundigheid blijken
- Ic Benadruk positieve kanten van jezelf

Bepaal schrijfdoelen (o.b.v. intentie en verwachting)
Kies Ia, Ib; doe Ic alleen voor gevraagde kenmerken

Hoe ga ik precies te werk?
Selecteren: Ga mogelijke inhouden na
- IIa Beschik over vereiste opleiding
- IIb Heb gevraagde kennis, vaardigheden en interesses
- IIc Ervaring is aanwezig, maar niet op alle fronten
- IId Ben ouder dan gevraagde leeftijd

Bepaal inhoud (o.b.v. schrijfdoelen Ib en Ic)
Kies IIa, IIb; noem van IIc alleen ervaring; zie af van IId

Ga mogelijke organisatiepatronen na
- IIIa Onderbouw (geschikt WANT argumenten)
- IIIb Concludeer (argumenten DUS geschikt)

Bepaal organisatie (o.b.v. schrijfdoel Ia)
Kies IIIa: geef eerst bewering en dan de argumenten

Vormgeven: Ga mogelijke vormgevingen van informatie na
- IVa Subject wel/niet vooropplaatsen
- IVb Wel/niet pronominaal verwijzen
- IVc Adverbiale bijzin wel/niet vooropplaatsen

Bepaal vormgeving (o.b.v. tekst-constraints en schrijfdoelen)
Kies ..... (verschilt per segment)

Tijdens het **vormgeven** bepaalt de schrijver een presentatiestijl voor de informatie. Hij bekijkt onder andere welke elementen hij voorop zal plaatsen en welk type verwijzing hij zal gebruiken. (Voor een didactische uitwerking van deze taalkundige verschijnselen zie Onrust, Verhagen & Doeve, 1993).

De sollicitant kan bijvoorbeeld een predikaat al dan niet vooropplaatsen. In (1a) bijvoorbeeld, wordt 'gevraagde ervaring' vooropgeplaatst waardoor dit aspect meer nadruk krijgt dan in (1b). Met het oog op schrijfdoel Ic zou (1a) de voorkeur hebben.

(1) a. De gevraagde ervaring heb ik in mijn huidige functie opgedaan.
b. In mijn huidige functie heb ik de gevraagde ervaring opgedaan.

Ook de manier waarop verwezen wordt naar bijvoorbeeld de ervaring moet worden bepaald. Als de ervaring nog niet eerder is genoemd in de tekst, zal een schrijver kiezen voor (2a). Is in het (direct) voorafgaande de ervaring al aangeduid, dan komen beknoptere verwijzingen in aanmerking, zoals (2b) en (2c). Bij het vormgeven zijn dus naast de schrijfdoelen ook constraints vanuit de reeds geproduceerde tekst van belang.

(2) a. Met het opzetten en verzorgen van taalcursussen voor volwassenen heb ik ruime ervaring.

b. Met het verzorgen van zulke taalcursussen heb ik ruime ervaring.
c. Daarmee heb ik ruime ervaring.

Begrijpen van informatie
Tijdens het begrijpen van informatie zoekt een lezer in iedere deelzin aanwijzingen (INSPECTEREN) om die deelzin met andere deelzinnen in verband te brengen (VERBINDEN). Dit gebeurt op basis van leesdoelen en koppeling aan wereldkennis (INTEGREREN). Het resultaat is een coherente mentale representatie van de tekst.

Stel dat iemand in een science-fiction-boek fragment (3) leest. Hoe begrijpt hij dan wat er beschreven wordt? Dat conceptuele proces wordt in tabel 2 geschetst.

(3) a. De bemanning betrad het platform
b en installeerde zich in de nauwe cabine.
c. De motoren maakten bij de start een oorverdovend lawaai.
d. Binnen een minuut waren ze uit het zicht verdwenen
e. en keerde de rust weer.

Om samenhang te ontdekken tussen de afzonderlijke deelzinnen van (3), worden ze **geïnspecteerd** op de aanwezigheid van linguïstische kenmerken. Deze kunnen aanwijzingen geven om de zinnen te **verbinden**. Zo duidt de samentrekking in (3b) op een relatie met (3a); het type werkwoord (actiewerkwoorden) en de verleden tijd geeft aan dat het waarschijnlijk om een reeks van gebeurtenissen gaat.[5]

Op basis van deze kenmerken kan een lezer wel samenhang tussen de afzonderlijke zinnen bepalen, maar daarmee is het fragment als geheel nog niet begrepen. De enige gemeenschappe-

Tabel 2. Het conceptuele proces bij het lezen van een tekstfragment uit een SF-boek.

Wat wil ik bereiken?	
Intentie:	Ik wil de tekst, zo compleet mogelijk, begrijpen
Verwachting:	Het zal wel over ruimtevaart gaan.
Hoe kan ik dat bij deze tekst realiseren?	
Integreren:	Ga mogelijke leesdoelen na
	Ia Bepaal samenhang tussen opeenvolgende zinnen
	Ib Zoek een gemeenschappelijke noemer (onderwerp, genre)
	Ic Relateer, waar mogelijk, tekst aan eigen wereldkennis
	Bepaal leesdoelen (o.b.v. intentie en verwachting)
	Kies Ia, Ic; stel voor Ib voorlopige invulling vast
Hoe ga ik precies te werk?	
Verbinden:	Ga mogelijke redenen voor verbinding na
	IIa Aanwezigheid van een connector
	IIb Semantisch type van hoofdwerkwoord
	IIc Aankaarten van een probleem/doel
	IId Positie van adverbiale bijzin
	Bepaal verbinding (o.b.v. leesdoelen)
	Kies ..... (verschilt per segment)
Inspecteren:	Ga mogelijke aanwijzingen voor verbinding na
	IIIa Zinsbouw: samentrekking, positie bijzin....
	IIIb Verwijzing: type en positie van anafoor....
	IIIc Woordbetekenis: actiewerkwoord, sequentie-aanduider..
	Bepaal aanwijzingen (o.b.v. tekst-constraints en leesdoelen)
	Kies ..... (verschilt per segment)

lijke noemer waar linguïstische aanwijzingen voor zijn, is het genre: narratief vanwege de reeks gebeurtenissen. Maar in welk kader die gebeurtenissen plaatsvinden, het onderwerp van de tekst, is niet duidelijk. Om het onderwerp te kunnen achterhalen is het noodzakelijk om de tekst, mede gestuurd door verwachtingen, te **integreren** met wereldkennis. Op grond van die kennis weet een lezer dat woorden als 'bemanning', 'platform', 'cabine', 'start' en 'oorverdovend lawaai van motoren' duiden op een raketlancering. Door de zinnen aan die kennis te relateren, komen ze in een betekenisvol perspectief te staan. Zo is duidelijk dat 'ze' in (3d) niet verwijst naar de dichtstbijzijnde mogelijke referent 'motoren' in (3c), en ook niet naar het enkelvoudige 'bemanning' in (3a), maar naar de niet genoemde raket met daarin de astronauten. Ook meerduidigheid van connectoren komt aan het licht door integratie met wereldkennis (Sanders, Spooren & Noordman, 1993). Zo is het 'en' in (3e) niet alleen sequentieel ('en toen'), maar ook causaal ('en daarom').

De karakterisering van de conceptuele componenten toont dat strategieën een belangrijke rol spelen. Een taalgebruiker stelt zichzelf doelen, denkt na over de realisering daarvan en bepaalt hoe hij dan precies te werk moet gaan. De rol van strategieën in de talige componenten is echter van een andere orde. Voor een betere kijk op dit verschil worden in de volgende paragraaf de conceptuele en talige processen met elkaar vergeleken.

2.3 Verschillen tussen conceptuele en talige processen

Het model van de taalgebruiker legt een vrij principiële scheidingslijn tussen de talige en de conceptuele processen (of, in meer traditionele termen, tussen de lagere, basale en de hogere, centrale processen). Hiermee wordt aangegeven dat deze processen geen *directe* invloed hebben op elkaars functioneren (Levelt, 1989, p. 103-105). Voorzover de conceptuele processen de talige beïnvloeden, gebeurt dat *indirect*. De VORMGEVING- en INSPECTEER-processen hebben toegang tot een arsenaal meta-linguïstische kennis en sturen daarmee formuleer- en ontleedbeslissingen (zie (1-3); voor inventarisaties van deze kennis, zie Sanders & Van Wijk, 1994a,b, en Van der Pool, 1995).

De scheiding tussen de conceptuele en talige processen komt geleidelijk tot stand tijdens het leren van de moedertaal (Wijnen, 1990). Vanaf de kleutertijd verloopt de verdere ontwikkeling van beide processen op een andere wijze.

De talige processen vertonen een geleidelijke uitbreiding van taalkennis (lexicon, grammatica, spelling e.d.) en een steeds meer routinematig gebruik van deze kennis. Deze ontwikkeling gaat geruime tijd door, maar bereikt op gegeven moment het punt 'materie meester'. De talige processen krijgen op den duur een zogeheten modulair karakter; ze worden ingekapseld en zijn ondoordringbaar voor andere procescomponenten.

De conceptuele processen maken eveneens een geleidelijke, cumulatieve ontwikkeling door, maar kunnen daarnaast ook reorganisaties ondergaan die in een nieuwe, diepere wijze van informatieverwerking resulteren. Deze dialectische veranderingen doen zich vooral voor in de REFLECTEER- en INTEGREER-processen (Bereiter & Scardamalia, 1987; Andriessen, 1991; Van der Pool, 1995). Op het conceptuele vlak kan men zich daarom steeds weer een beginner voelen, zij het op een hoger niveau van functioneren. De conceptuele processen krijgen een meer intentioneel karakter; ze worden een geïntegreerd onderdeel van een probleem-oplossend denkproces.

Het verschillend karakter van de talige en conceptuele processen bepaalt het criterium voor vaardigheid. Bij de talige processen is dit automatisering: de snelle en accurate uitvoering van context-vrije procedures. Bij de conceptuele processen is dit echter intentionaliteit: de doelgerichte en gecoördineerde uitvoering van context-gevoelige strategieën.

De automatisering van de talige processen is van groot belang omdat deze processen anders ten koste van de conceptuele processen een beroep blijven doen op verwerkingscapaciteit (Wel-

schen, 1982; Van Dijk & Kintsch, 1983; Bossers, 1993a,b). De gevolgen van deze aandachtscompetitie treden ook op wanneer de taalgebruiker zelf besluit om de aandacht te verdelen over talige en conceptuele aspecten zoals het geval is bij revisie- en correctiewerk (Broekkamp & Van den Bergh, dit nummer).

Bij de conceptuele processen zijn strategieën een essentieel bestanddeel. Vaardigheid wordt hier bepaald door het strategie-repertoire en de daarmee samenhangende meta-cognitie ('self-awareness'; Flower, 1989, p. 45). Bij de talige processen spelen strategieën een geheel andere rol. Van het normale functioneren maken zij geen deel uit (Van Dijk & Kintsch, 1983, p. 24; Bossers, 1993a,b). Alleen wanneer de processen moeizaam verlopen door een gebrekkige taalbeheersing of een fysiek trauma (beroerte e.d.), kan de taalgebruiker terugvallen op noodoplossingen in de vorm van compensatie-strategieën. Voorbeelden zijn de vermijdingsstrategieën in de adaptatietheorie van agrammatisme (Kolk, 1987) en de raadstrategieën in de methode van het Voorspellend Lezen (Westhoff, 1993; Hacquebord, dit nummer).

Met deze functionele dispariteit is nog niet alles gezegd over de rol van strategieën in de mentale processen van de taalgebruiker. Maar om meer in detail te kunnen treden, moet er eerst een duidelijk zicht zijn op de betekenis van het begrip *strategie*. Aan die verheldering is de volgende paragraaf gewijd.

3. STRATEGIEËN: EEN BEGRIPSBEPALING

In (4) staat een karakterisering van het begrip *strategie* die nauw aansluit bij het gangbare idee (zie bijv. Van Dijk & Kintsch, 1983, p. 62-68). Als volzin luidt de definitie: een strategie is de globale representatie van de procedure(s) om een bepaald doel te realiseren. Oftewel, een strategie is het van een handelingsschema voorziene oplossingsvoorstel voor een probleem.[6]

(4) DOEL	->	Wat wil je bereiken?	(probleem)
Strategie	->	Hoe kun je dat realiseren?	(oplossing)
procedure	->	Hoe ga je precies te werk?	(handeling)

Aan de hand van deze definitie kunnen vier cruciale kenmerken geadstrueerd worden. De eerste volgt er zelfs direct uit.

A. Een strategie is onlosmakelijk verbonden aan een motiverend doel en één of meer concrete procedures, maar kan niet aan een van beide gelijkgesteld worden.
B. De relaties doel-strategie en strategie-procedure zijn nooit logisch-deductief, maar altijd instrumenteel.

Een bepaald doel kan vrijwel altijd langs meer dan één weg gerealiseerd worden. In (5) staat hiervan een voorbeeld: er zijn diverse strategieën om van je honger af te komen, en iedere strategie biedt de keus uit een aantal procedures.

(5) DOEL	Ik wil van mijn honger af
Strategie 1	Zorg ervoor dat je maag gevuld wordt
procedure 1a	Ga naar de ijskast en plunder hem
procedure 1b	Snoep verstandig, neem een appel
Strategie 2	Zorg ervoor dat je afgeleid wordt
procedure 2a	Zet tv aan en begin een zap-soap-search
procedure 2b	Ga de goudviskom schoonmaken
Strategie 3	Zorg ervoor dat het gevoel verdwijnt
procedure 3a	Drink een glas koud water
procedure 3b	Steek een sigaret op

C. Een strategie maakt in de regel deel uit van een netwerk met zowel een lineaire als een hiërarchische organisatie.

In (5) staat een voorbeeld van een louter lineaire opzet. In (6) komt ook het hiërarchische aspect aan bod. De objecten van procedure 1a zijn direct herkenbaar en de handelingen dus even direct uitvoerbaar, die van 1b echter niet. De uitvoering van procedure 1b vereist daarom de activering van een aantal subdoelen en hun bijbehorende strategieën.

(6) DOEL	Ik wil weten waar deze tekst globaal over gaat
Strategie 1	Loop de tekst na op thema-indicaties
procedure 1a	Kijk naar titel, tussenkopjes en illustraties (Bimmel & Westhoff, 1993)
procedure 1b	Kijk naar topic-zinnen en concluderende zinnen (Onrust, Verhagen & Doeve, 1993)
SUBDOEL 1b1	Ik wil weten wat de topic-zinnen zijn
Strategie 1b1	Loop de voor deze zin geëigende plaatsen langs
procedure 1b1a	Kijk naar de beginzin van iedere alinea (MacDonald, 1983)
SUBDOEL 1b2	Ik wil weten wat de concluderende zinnen zijn
Strategie 1b2	Loop de voor deze zin geëigende plaatsen langs
procedure 1b2a	Kijk naar de slotzin van iedere alinea (MacDonald, 1983)

D. Het adequaat gebruik van strategieën vereist een tactiek.

Tactiek is het organiseren van strategieën met het oog op bepaalde omstandigheden. Op de eerste plaats is tactiek een kwestie van afstemming op de situatie: het doel wordt in een context geplaatst. Bijvoorbeeld, in (5) valt strategie 2 af, als je aan het studeren bent en nog niet wilt stoppen.

Op de tweede plaats is tactiek een kwestie van kosten/baten-analyse. Aan het doel wordt een optimaliseringscriterium gehecht: hoe direct, goed, snel, zuinig etc. mogelijk wil je het bereiken? Bijvoorbeeld, in (5) vormen de procedures 1b en 3a de optimale keuze-ruimte voor wie, op termijn, gezond wil blijven; de procedures 1a en 3b hebben de voorkeur van degene die directe bevrediging van fysieke lust hoger waardeert.

Bij de invulling van het taalonderwijs moeten deze vier kenmerken nauwlettend in het oog worden gehouden. Vooral een gelijkstelling van strategieën aan procedures en het werken in een tactisch vacuüm kunnen ongelukkig uitpakken. De psycholinguïstische argumenten hiervoor komen in de volgende paragraaf aan de orde.

4. DE ONTWIKKELING VAN CONCEPTUELE VAARDIGHEDEN

In paragraaf 2.2 en 2.3 zijn de conceptuele processen benaderd in hun 'expert-like' uitvoering, als de doelbewuste en gecoördineerde uitvoering van strategieën[7]. In het alledaagse gebruik hebben conceptuele processen echter veel meer met de talige processen gemeen: ze putten uit een arsenaal aan vuistregels en passen die routineus toe. Want voor de hogere denkprocessen geldt evenzeer de simpele leefregel van cognitieve zuinigheid (Noordman & Vonk, 1987). Waarom moeilijk doen, als je gewoon ook al een heel eind kunt komen?

Deze kwalitatieve verschillen in het functioneren van de conceptuele component hangen samen met het onderscheid tussen spontaan verworven en aangeleerde vaardigheden (verg. 'natural versus problematic' bij Bereiter & Scardamalia, 1987, p. 4). Bij allerlei sensorische, motorische, sociale en intellectuele vaardigheden is voor normaal functioneren een zekere graad

van beheersing gewenst die verkregen kan worden door langs natuurlijke weg ervaring op te doen (in spel, door 'modelling' e.d.). Deze spontaan verworven vaardigheden worden in principe door iedere taalgebruiker beheerst. Daarnaast zijn er vaardigheden die slechts in bepaalde situaties beheerst dienen te worden, en die ook alleen door specifieke studie of training, vaak moeizaam, verworven kunnen worden. Deze aangeleerde vaardigheden beheerst men als men daartoe aangezet wordt, gemotiveerd is en natuurlijk de capaciteiten heeft. Bijvoorbeeld, iedereen leert lopen en praten. Heel wat minder individuen slagen er echter in om op niveau een marathon af te leggen, een tango te dansen, een voordracht te houden of een aria te zingen.

De ontwikkeling van de talige component voltrekt zich grotendeels spontaan, d.w.z. op basis van alledaagse contacten in interactief taalverkeer (Wijnen, 1990). Als men naar school gaat en leert schrijven en lezen, betekent dat een confrontatie met nieuwe senso-motorische taken en een forse groei van de mentale kennisbestanden (lexicon, wereldkennis). Het verloop van de processen worden er echter niet wezenlijk anders door.

Ook de conceptuele component ontwikkelt zich eerst zonder 'school te gaan'. Deze spontaan verworven vaardigheid steunt echter op twee pijlers. Ten eerste blijft het proces heel dicht bij het direct voor handen zijnde materiaal: de letterlijke tekst bij verwerking en de reeds aanwezige kennis bij produktie. Er wordt weinig aan toegevoegd door eigen cognitieve activiteit (een tekstplan, een macrostructuur etc.). En ten tweede is het proces sterk afhankelijk van de inbreng van een conversatie-partner. In gesprekken heeft die conversatie-partner vele (non)verbale feedback-mogelijkheden om als aandrijfkracht èn als monitor te fungeren.

Aan deze dubbele afhankelijkheid worden met de jaren steeds meer grenzen gesteld. Een principiële oorzaak is de verschuiving van de prioriteit bij de conceptuele vaardigheden: naast adequaat meedoen aan alledaagse taalverkeersituaties moet men ook kritisch en zelfstandig kunnen denken. Van de taalgebruiker wordt in toenemende mate een initiërende en reflectieve houding gevraagd.

Een praktische oorzaak is dat bij lezen en schrijven het contact met een interactieve partner verloren gaat. De 'in vivo'-gebruikssituatie wordt een 'in-vitro'-werkomgeving (Van Wijk, 1989). De taalgebruiker moet voortaan zelf het proces opstarten, gaande houden en bijsturen.

De conceptuele ontwikkeling mondt (in het gunstige geval) uit in de opbouw van een geheel nieuwe vorm van informatieverwerking. Bij het lezen is dit bekend als 'deep processing' (zie bijv. Schank & Lebowitz, 1980), bij het schrijven als 'knowledge transforming' (Bereiter & Scardamalia, 1987). In Tabel 3 worden de aangeleerde en spontaan verworven procesuitvoeringen met elkaar vergeleken.

Het spontane proces is procedureel van aard; het verloopt aan de hand van vuistregels waarin door ooit gevormde associaties patronen zijn ontstaan. Daarbij gaat men vrijwel geheel uit van de eigen wereldkennis en de letterlijke informatie in de tekst. Het gevolg is een vrij stereotiepe omgang met informatie die sterk gericht is op assimilatie ('het nieuwe bij het bekende onderbrengen').

Het aangeleerde proces is strategisch van aard: het verloopt in de vorm van een redeneerproces waarin zelf gestelde doelen als richtsnoer fungeren. Daardoor ontstaat een flexibele informatieverwerking die de mogelijkheid creëert om tot nieuwe inzichten te komen (accommodatie).

Tabel 3. Karakterisering van de spontane en de aangeleerde uitvoering van conceptuele processen.

Spontaan	Aangeleerd
Taak uitvoerend:	Probleem oplossend:
Procedureel	Strategisch
associatief	redenerend
data-driven	goal-driven
stereotype	flexibel
assimilerend	accommoderend

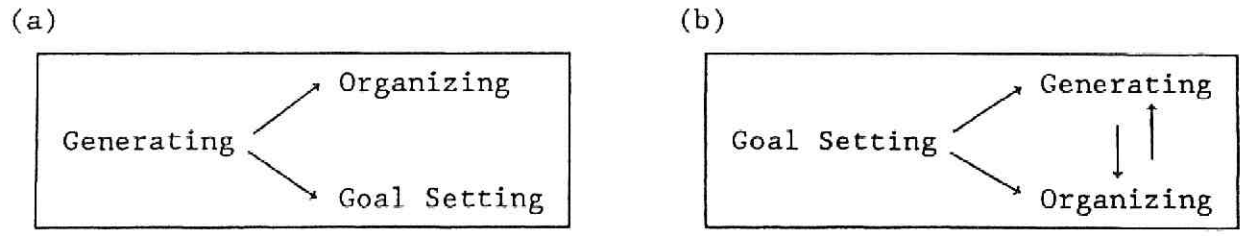


Fig. 2. Conceptuele processen tijdens het schrijven bij (a) een procedurele werkwijze (Hayes & Flower, 1980) en (b) een strategische aanpak (Bereiter & Scardamalia, 1987).

Het verschil tussen de procedurele en de strategische aanpak wordt in Figuur 2 geïllustreerd voor de conceptuele component van het schrijven. De a-versie is afkomstig uit het model van Hayes en Flower (1980). In weerwil van hun pretentie de expert te modelleren, geeft hun voorstel het spontaan verworven proces weer. De b-versie is Bereiter en Scardamalia's (1987) weergave van het aangeleerde proces.

Bij de procedurele aanpak in (a) begint het schrijfproces door uit de intentie (opdracht) een onderwerp (topic) te lichten en daaromheen een netwerk van associaties te bouwen. Dit levert de steekwoorden op (data) waarmee in het geheugen naar aanvullende informatie wordt gezocht. Uit de informatie die zo beschikbaar komt, selecteert de schrijver de bruikbare elementen en plaatst ze in een structuur. Uit het geheugen kunnen ook criteria opkomen om de tekst te beoordelen. Deze worden onthouden als doelen voor de tekstrevisie, dus niet voor de inhoudsgeneratie zelf. Deze stereotiepe en reactieve werkwijze put alleen uit reeds beschikbare wereldkennis (assimilatie oftewel 'knowledge telling').

Bij de strategische aanpak in (b) begint het proces met een redeneergang waarin de intentie (opdracht) wordt geanalyseerd en gepreciseerd. Daaruit worden doelen afgeleid die als richtsnoer fungeren voor de inhoudselectie en de tekstorganisatie. Deze twee processen staan onderling in nauw contact. Het beschikbaar komen van bepaalde informatie stelt de schrijver voor de taak die ergens in te passen. De selectie van een bepaalde retorische vorm dwingt de schrijver te zoeken naar een plausibele invulling. Zo'n flexibele en kritische werkwijze biedt de mogelijkheid om tot nieuwe inzichten en kennis te komen (accommodatie oftewel 'knowledge transforming').

Het procedureel ingevulde schrijfproces beperkt zich dus tot een eenzijdig op inhoud georiënteerde opzet van het SELECTEER-proces waardoor de taak verwordt tot (associatieve) 'memory-retrieval'. Het strategisch uitgevoerde proces plaatst een REFLECTEER-proces boven de SELECTEER-operaties. Het laat binnen SELECTEER een actieve wisselwerking tussen deelprocessen plaatsvinden. De taak wordt een kwestie van (intentionele) 'problem solving'[8].

5. STRATEGIEËN EN TAALONDERWIJS

De voorgaande schets van de conceptuele ontwikkeling laat zien dat strategieën de kern uitmaken van de 'geschoolde' produktie en verwerking van tekst. Om zover te kunnen komen, moet het de leerling mogelijk worden gemaakt om tot een actieve en reflectieve werkhouding te komen. Het taalonderwijs is daarvoor een aangewezen plaats. Door een sterk accent te leggen op strategieën, meta-cognitie en zelfregulatie zijn de belangrijkste ingrediënten aanwezig van proces-gericht onderwijs. Daarbij moeten echter drie punten niet aan de aandacht ontsnappen.

Onderwijs in strategieën vergt een gedegen empirisch en theoretisch fundament. Je kunt niet te veel vertrouwen op plausibiliteitsargumenten ontleend aan gezond verstand en praktische ervaringen. Gebeurt dat wel, dan ontstaat er maar al te gemakkelijk een fragmentarische presentatie van leerstof. In (6) zagen we hoe uit verschillende bronnen deeloplossingen werden aangedragen. En dat het eigen voorstellingsvermogen maar al te snel te kort schiet, bleek bij de

bespreking van Hayes en Flower's modellering van het conceptuele proces (zie paragraaf 4; Bereiter & Scardamalia, 1987, p. 25; Vos, 1985; Van der Geest, Schellens & Van Waes, 1992).

Strategisch werken is een kwestie van goed weten waarmee je bezig bent (meta-cognitie) en van procesmanagement (zelfregulatie). Door het feedback-isolement bij het schrijven en lezen wordt de eigen behendigheid in de processturing van steeds meer belang. Hoe ga je procedures oproepen, starten, uitvoeren en beëindigen als onderdeel van een strategie?

Deze controlerende taak maakt de conceptuele processen op zich al inspannend. Daarbij komt dan ook nog de uitvoerende last: bij iedere strategie moeten de op dat moment geëigende procedures worden bepaald. De verlokking is dan ook groot om het intentionele te mijden en je te beperken tot het methodische. De strategie wordt ingewisseld voor een procedure. En strategisch werken verwordt dan tot het putten uit een trucendoos. Bijvoorbeeld, als je in (6) procedure 1a voor de strategie aanziet, dan heb je een vuistregel geleerd in plaats van een strategie (voor een ander sprekend voorbeeld, zie Bereiter & Scardamalia, 1987, p. 322).

Strategisch werken betekent niet het hebben van pasklare antwoorden, maar het flexibel uitvoeren van een afwegingsproces. De notie *tactiek* vormt hierin een onmisbaar onderdeel. Strategieën leer je op basis van redenering en vergelijking, dus niet in isolement, maar in samenhang met elkaar.

Tactisch werken staat gelijk aan de ontwikkeling van REFLECTEER- en INTEGREER-vaardigheden. Het zet aan tot een meer flexibele en diepergravende wijze van informatieverwerking. Strategieën zijn niet louter hulpmiddelen om taken hanteerbaar te maken. Zij functioneren ook als uitdagingen om een taak op een hoger niveau uit te voeren. Strategieën kunnen fungeren als 'change inducing agents'. Vooral deze functie zou centraal moeten staan in procesgericht taalonderwijs. Het is echter zeer aanlokkelijk om de meeste aandacht naar SELECTEER/VORMGEVING- en INSPECTEER/ VERBIND-strategieën uit te laten gaan: het werk blijft daardoor nogal concreet, vergt geen overdosis denkkracht en levert op korte termijn zelfs tastbaar resultaat op. Strategisch denken wordt dan vervangen door procedureel uitvoeren. Op den duur blijkt pas dat dat een zelf opgezette val is (voor een zeer kritische stellingname, zie Bereiter & Scardamalia, 1987, p. 186-189).

6. BESLUIT

Proces-gericht taalonderwijs bezit van nature een grote mate van validiteit: taalvaardigheden worden vanuit hun dynamiek en in samenhang met elkaar benaderd. In taalbeheersingsonderzoek zijn daartoe reeds vruchtbare aanzetten gegeven; in het taalonderwijs is van een fundamentele heroriëntatie helaas nog weinig sprake (zie Rijlaarsdam, Oostdam & Bimmel, dit nummer). Een ecologisch valide vertrekpunt is echter geen garantie voor een even valide uitwerking. Strategieën en procedures moeten duidelijk gescheiden blijven omdat zij zowel begripsmatig als procesmatig niet aan elkaar gelijk gesteld kunnen worden. Intentioneel werken is iets anders dan methodisch werken. Als strategie-onderwijs wordt ingeëngd tot het leren van procedures, dan bereikt men bij de leerling wel een verrijking van de basale (spontane) vorm van informatieverwerking, maar benut men niet de mogelijkheid om meer complexe (aangeleerde) vormen van informatieverwerking te ontwikkelen.[9]

Het hier uitgewerkte model is een aanzet tot theoretische onderbouwing van proces-gericht taalonderwijs. Om processen en strategieën een meer centrale plaats te kunnen geven in het onderwijs is het echter nodig om naast de conceptualiserende functie in lezen en schrijven (taalgebruiksstrategieën) ook aandacht te besteden aan de regulerende functie (controle en leerstrategieën). De psycholinguïstiek heeft altijd beduidend meer oog gehad voor de conceptualiserende kant, de onderwijskunde daarentegen voor de regulerende. Voor een echt algemeen model van de taalgebruiker zal tussen beide disciplines een brug moeten worden geslagen. De auteurs van dit themanummer hopen daaraan een bijdrage te hebben geleverd.

NOTEN

1. Voor algemene modellen van de taalgebruiker, zie Dijkstra & Kempen (1993) en Levelt (1989). Voor specifieke modellen van produktie of verwerking van geschreven tekst, zie Bereiter & Scardamalia (1987), Flower & Hayes (1981), Cooper & Matshuhashi (1983), Witte & Cherry (1986), McCutchen & Perfetti (1982), Van Dijk & Kintsch (1983), Balota, Flores d'Arcais & Rayner (1990), Rayner & Pollatsek (1989) en Just & Carpenter (1987).
2. Overal waar in deze tekst mannelijke woorden zijn gebruikt, kan ook de vrouwelijke vorm worden gelezen. Voor de leesbaarheid is ervoor gekozen alleen mannelijke verwijzingen te gebruiken.
3. Uitbreiding van het lexicon verloopt eveneens op deze manier. Voor de volledigheid moet worden opgemerkt dat er dus ook pijlen vanuit de conceptuele componenten naar het lexicon dienen te staan. Deze pijlen zouden dan echter wel 'ontwikkeling' symboliseren in tegenstelling tot de andere pijlen die 'gebruikmaking-van' of 'doorsturen' symboliseren.
4. Reflectie vervult bij de evaluatie de functie van 'monitoring', het geven van feedback op het conceptuele deel van het produktieproces. (Zie Levelt 1989, p. 468 e.v. voor een beknopt overzicht van de discussie over de rol en de plaats van monitoring.)
5. Voor een verdere uitwerking van deze aanwijzingen voor verbindingen tussen zinnen zie Sanders & Van Wijk (1994a, b) en Van der Pool (1995).
6. Min of meer synoniem met *strategie* zijn de termen *heuristiek* en *beleid*, en met *procedure* de termen *vuistregel* en *implementatie*. De combinatie van *doel* en *strategie* wordt aangegeven met de term *plan*, de combinatie van *strategie* en *procedure* met de termen *werkwijze* en *methode*.
7. De tabellen 1 en 2 zijn bijvoorbeeld uitwerkingen van de strategie karakterisering in (4).
8. Deze kwalitatieve verschillen in de conceptuele component zijn al langs verschillende ingangen empirisch verkend: zie voor tekstplanning Andriessen (1991, p. 199), voor schrijven Van der Pool (1992), voor tekstrevisie Sanders en Van Wijk (1994b) en voor literair lezen Andringa (dit nummer).
9. De geschiedenis leert dat, ondanks alle goede voornemens, het gevaar van zo'n inenging verre van irreëel is. In de taalbeheersing van de tachtiger jaren verschoof bijvoorbeeld de aandacht van 'geen fouten maken' naar 'doelmatig communiceren', van de conventionaliteit van taalgedrag naar de functionaliteit ervan. In allerlei zogeheten peilingsonderzoeken zijn bij leerlingen hun functionele taalvaardigheden bepaald. In deze onderzoeksprojecten waren zowel de taken als de evaluatie-criteria grotendeels op gezond verstand gebaseerd, deels door de eigen neiging om de praktijk vooruit te laten lopen op de theorie, deels door de eis van subsidiegevers om snel met concrete resultaten te komen. Met alle gevolgen vandien. Analyses die in een later stadium op een deel van het materiaal zijn uitgevoerd, laten zien dat het taalgedrag voor een belangrijk deel niet op functionele maar op conventionele criteria is gewaardeerd (Van Wijk, 1989; Boekelder, De Glopper & Van Wijk, 1991; Tel, De Glopper & Van Wijk, 1994).

LITERATUUR

Anderson, R. & Pichert, J. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17*, 1-12.

Andriessen, J. (1991). *Minimal strategies for coherent text production.* Proefschrift, Rijksuniversiteit Utrecht.

Balota, D., Flores d'Arcais, G. & Rayner, K. (1990). *Comprehension processes in reading*. Hillsdale: Erlbaum.

Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale: Erlbaum.

Bergh, H. van den & Meuffels, B. (1993). Schrijfvaardigheid. In A. Braet & J. van de Gein (red.). *Taalbeheersing als tekstwetenschap: terreinen en trends*. (p. 177-211). Dordrecht: ICG.

Bergh, van den H., Rijlaarsdam, G. & Breetvelt, I. (1992). Het stellen van doelen gedurende het schrijfproces: Relaties met produktkwaliteit. *Tijdschrift voor Taalbeheersing, 14*, 221-233.

Bimmel, P. & Westhoff, G. (1993). Leesstrategieën: training in het moedertaalonderwijs, transfer naar het vreemde-talenonderwijs. *Levende Talen, nr. 485*, 577-582.

Boekelder, A., de Glopper, K. & van Wijk, C. (1991). Tekstanalytisch onderzoek van een functionele schrijfopdracht: de inhoud, opbouw en beoordeling van excuusbrieven. *Toegepaste Taalwetenschap in Artikelen, 40*, 23-34.

Bossers, B. (1993a). VT-lezen: leesstrategieën of taalkennis? *Levende Talen, nr. 478*, 133-139.

Bossers, B. (1993b). Nogmaals: leesstrategieën en VT-kennis. *Levende Talen, nr. 481*, 326-330.

Bower, G. & Cirilo, R. (1985). Cognitive psychology and text processing. In T. van Dijk (ed.) *Handbook of discourse analysis (vol. 1)*. (p. 71-105). Orlando: Academic Press.

Braet, A. & Janssen, J. (1993). Het taalvaardigheidsonderwijs. In A. Braet & J. van de Gein (red.). *Taalbeheersing als tekstwetenschap: terreinen en trends*. (p. 213-237). Dordrecht: ICG.

Commissie Herziening Eindtermen (1990). *Kerndoelen voor de basisvorming in basisonderwijs en voortgezet onderwijs*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.

Cooper, C. & Matshuhashi, A. (1983). A theory of the writing process. In M. Martlew (ed.). *The psychology of written language: a developmental approach*. (p. 3-39). London: Wiley.

Dijk, T. van & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Dijkstra, A. & Kempen, G. (1993). *Taalpsychologie*. Groningen: Wolters-Noordhoff.

Flower, L. (1989). *Problem solving strategies for writing* (third revised edition). Orlando: Harcourt Brace Jovanovich.

Flower, L. & Hayes, J. (1981). Plans that guide the composing process. In C. Frederiksen & J. Dominic (eds.). *Writing; the nature, development, and teaching of written communication. Vol. 2: Writing: process, development and communication* (p. 39-58). Hillsdale: Erlbaum.

Geest, T. van der, Schellens, P. & van Waes, L. (1992). Het model van Flower en Hayes: een cognitief procesmodel of een retorisch taakmodel? *Tijdschrift voor Taalbeheersing 14*, 161-176.

Hayes, J. & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (eds.). *Cognitive processes in writing*. (p. 3-30). Hillsdale: Erlbaum.

Just, M. & Carpenter, P. (1987). *The psychology of reading and language comprehension*. Boston: Allyn and Bacon.

Kolk, H. (1987). A theory of grammatical impairment in aphasia. In G. Kempen (ed.). *Natural Language Generation*. (p. 377-391). Dordrecht: Nijhoff.

Levelt, W. (1989). *Speaking; from intention to articulation*. Cambridge, MA: MIT Press.

MacDonald, N. (1983). The UNIX Writer's Workbench Software: Rationale and Design. *The Bell System Technical Journal, 62*, 1891-1908.

Matsuhashi, A. (1981). Pausing and planning: the tempo of written discourse production. *Research in the Teaching of English, 15*, 113-134.

McCutchen, D. & Perfetti, C. (1982). Coherence and connectedness in the development of discourse production. *Text, 2*, 113-139.

Noordman, L. & Vonk, W. (1987). Selectieve verwerking van tekst. *Tijdschrift voor Taal- en Tekstwetenschap, 7*, 57-69.

Onrust, M., Verhagen, A. & Doeve, R. (1993). *Formuleren*. Houten: Bohn Stafleu Van Loghum.

Pool, E. van der (1995). *Conceptual processes in writing. A text-analytical study of developmental aspects*. Proefschrift, K.U. Brabant.

Pool, E. van der (1992). Planning and organisation in descriptive texts; a text-analytical study. *Proceedings of the CLS Ph.D. Conference 1992* (p. 43-52). Tilburg/Nijmegen: Centre for Language Studies.

Propp, V. (1984). *Theory and history of folklore*. Manchester: Manchester University Press (Vertaald door A. Martin, R. Martin e.a. geredigeerd met een introductie en noten van A. Liberman).

Rayner, K. & Pollatsek, K. (1989). *The psychology of reading*. Englewood Cliffs: Prentice-Hall.

Sanders, T., Spooren, W. & Noordman, L. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics, 4*, 93-133.

Sanders, T. & van Wijk, C. (1994a). Procedures voor Incrementele Structuuranalyse (PISA): uitleg en toepassing op verklarende teksten. In A. Maes, P. van Hauwermeiren & L. van Waes (red.). *Perspectieven in taalbeheersingsonderzoek*. (p. 191-203). Dordrecht: ICG.

Sanders, T. & van Wijk, C. (1994b). PISA – A procedure for analyzing the structure of explanatory texts. Manuscript, RU Utrecht/ KU Brabant (geaccepteerd door *Text*).

Schank, R. & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale: Erlbaum.

Schank, R. & Lebowitz, M. (1980). Levels of understanding in computers and people. *Poetics, 9*, 251-273.

Tel, M., de Glopper, K. & van Wijk, C. (1994). De sollicitatiebrief als functionele schrijfopdracht. In A. Maes, P. van Hauwermeiren & L. van Waes (red.). *Perspectieven in taalbeheersingsonderzoek*. (p.216-226). Dordrecht: ICG.

Vos, J. (1985). Schrijfprocesonderzoek en schrijfmethoden: het onderzoek van Flower en Hayes en het leerboek van Flower. *Tijdschrift voor Taalbeheersing, 7*, 258-269.

Welschen, A. (1982). Formuleervaardigheid en de cognitieve balans bij het schrijven. *Tijdschrift voor Taalbeheersing, 4*, 131-162.

Westhoff, G. (1993). Onderwijs in leesstrategieën: theoretische achtergrond, praktisch nut. *Levende Talen, nr. 480*, 266-270.

Wijnen, F. (1990), *On the Development of Language Production Mechanisms*. Proefschrift, Katholieke Universiteit Nijmegen.

Wijk, C. van (1989). Taal in vivo; een reactie op Henk Blok's 'Taal voor alledag'. *Gramma, tijdschrift voor taalkunde, 13*, 57-74.

Wijk, C. van (1992). Information analysis in written discourse. In J. de Jong & L. Verhoeven (eds.). *The construct of language proficiency*. (p. 85-99). Amsterdam: Benjamins.

Witte, S. & Cherry, R. (1986). Writing processes and written products in composition research. In C. Cooper & S. Greenbaum (eds.) *Studying writing; linguistic approaches*. (p. 112-153). Beverly Hills: Sage.

Manucript ontvangen 16-5-1994
Definitieve versie ontvangen 22-11-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 3, pp. 215-228

Strategische competentie in het talenonderwijs

Gert Rijlaarsdam[1], Ron Oostdam[2] en Peter Bimmel[3]
[1]*Instituut voor de Lerarenopleiding (ILO), Universiteit van Amsterdam**
[2]*SCO/Kohnstamm Instituut, Faculteit der Pedagogische en Onderwijskundige Wetenschappen, Universiteit van Amsterdam*
[3]*Interfacultair Instituut voor Lerarenopleiding, Onderwijsontwikkeling en Studievaardigheden (IVLOS), Universiteit Utrecht*

ABSTRACT

The problematic nature of coordinating mother tongue and foreign language education is widely acknowledged. Recently, attention has shifted to language problem solving strategies, metacognition and selfregulation as factors for facilitating transfer. In this article, these factors are incorporated in a model for language processing and learning. The model consists of three major components: the input, the long term memory and the working memory. This model has been developed drawing upon recent insights of literature on language learning, language processing and transfer. A multidimensional classification scheme, based on this model, has been developed for analyzing a representative sample of exercises in some of the most widely used textbooks for Dutch (mother tongue), English and French (as foreign languages). Results of this comparative analysis show that textbooks for mother tongue education hardly use the potentialities for transfer of language processing and language learning strategies. Foreign language textbooks lack almost any opportunity for strategic language learning.

1. INLEIDING

In de vakdidactische literatuur op het gebied van talenonderwijs is veel aandacht voor taalgebruiks- en taalleerstrategieën. Dit geldt niet alleen voor het moedertaalonderwijs, maar ook voor het vreemde- en tweede-taalonderwijs (cf. Bialystok 1990, Bimmel 1992, O'Malley & Chamot 1990, Faerch & Kasper 1980, 1986, Van Hout-Wolters 1992, Oostdam & Rijlaarsdam 1993 a & b, Oxford 1990, Poulisse 1986, Poulisse, Bongaerts & Kellerman 1987, Schouten-van Parreren 1985, Wenden & Rubin 1987, Westhoff 1981, Willems 1986). Deze aandacht voor strategieën heeft uiteindelijk een officieel beslag gekregen in de kerndoelen voor de Basisvorming (Commissie Herziening Eindtermen 1990). Het (talen)onderwijs zou meer procesgericht moeten worden: niet alleen aandacht geven aan wat leerlingen doen, maar vooral ook aan hoe ze het doen.

Bij het verwerven van strategieën speelt zelfregulatie een belangrijke rol. Onder zelfregulatie wordt in de leerpsychologie verstaan het kunnen plannen, bewaken ('monitoren') en evalueren van de uitvoering van taken. Wanneer het taal betreft gaat het om het reguleren van het eigen taalgebruiks- en taalleerproces (cf. Simons 1988, Bimmel 1993). Noodzakelijke voorwaarde voor zelfregulatie is in de eerste plaats kennis over eigen strategieën: weten uit welke componenten die strategieën zijn opgebouwd ('knowing that') en weten wanneer zij wel of niet kunnen worden ingezet ('knowing when'). Metacognitieve kennis over de eigen taalgebruiks- en taalleerstrategieën is een tweede voorwaarde. Dergelijke kennis omtrent de eigen strategieën bestaat er ondermeer uit dat leerlingen weten welke strategie voorhanden is en hoe goed zij zijn in het toepassen van de strategie.

Reflecteren op (eigen) strategisch handelen is binnen het onderwijsleerproces een middel om zelfregulatie te bewerkstelligen. Reflectieactiviteiten hebben als doel ervoor te zorgen dat leerlingen een metacognitieve ervaring opdoen, een bewustwording van de wijze waarop zij een taak hebben uitgevoerd (cf. Flavell 1978). Bij doelgerichte reflectie wordt aan leerlingen ge-

*Adres: Herengracht 256, 1016 BV Amsterdam.

vraagd te verwoorden hoe zij tot een oplossing zijn gekomen: Wat was je doel? Welke stappen heb je gezet? Wat was het resultaat? Hoe verklaar je dat? (Bimmel & Westhoff 1993). Verondersteld mag worden dat training in zelfregulatie en het opbouwen van metacognitieve kennis, de beschikbaarheid, bruikbaarheid en daarmee de wendbaarheid van taalgebruiks- en leerstrategieën vergroot (cf. Van Parreren 1981, Simons 1990).

De didactiek van dergelijk strategie-onderwijs heeft het (leer)strategisch handelen van de leerling als object. Enkele hoofdkenmerken van deze didactiek kunnen worden ontleend aan de leerpsychologische literatuur over transfer (Larkin 1989, Salomon & Globerson 1987, Salomon & Perkins 1989, Simons 1990). Hierin wordt het belang onderstreept van expliciete leerdoelen (het aangeven waarom iets geleerd moet worden door onder andere te verwijzen naar toepassingscontexten). Daarnaast verhogen authentieke en betekenisvolle leer- of taaltaken[1] de kans dat leerlingen inzien waarom een strategie nuttig en zinvol is ('meaningful learning'). Daarbij is het met name bij 'near transfer' raadzaam zorg te dragen voor verschillende, uiteenlopende toepassingscontexten met het oog op decontextualisatie ('low road transfer'; Salomon & Perkins 1989). Wil er transfer optreden dan moet de leerder overeenkomsten tussen een nieuwe taak en een eerder geleerde taak waarnemen en begrijpen. Voor 'low road transfer' wordt bovendien het inbouwen van pregnante evocatoren in leertaken aanbevolen (Van Parreren 1974). Dat betekent dat ervoor gezorgd moet worden dat een getrainde strategie duidelijk geassocieerd wordt met eigenaardigheden van mogelijke toepassingscontexten. Dergelijke eigenaardigheden evoceren strategiegebruik in latere toepassingscontexten.

Gaat het om 'far transfer' dan is 'mindful abstraction' de aangewezen maatregel: via welbewuste reflectie komen tot een generalisatie van een probleemtype en de gehanteerde oplossingsstrategie ('high road transfer'). De leerder moet eerst een leertaak begrijpen met zijn interne en externe functionele/structurele relaties. Vervolgens moet deze hele taak- en oplossingsstructuur naar een abstracter niveau getild worden en zelf ook weer in het lange-termijngeheugen opgeslagen worden (cf. Simons 1990).

In dit artikel bespreken we een model voor het gebruiken en leren van taal (Oostdam & Rijlaarsdam 1995), waarin de begrippen zelfregulatie, strategische competentie en metacognitie centraal staan. Het model bouwt voort op twee theoretische kaders: de talige competentie wordt verdeeld in een aantal deelcompetenties (cf. Bachman 1990, Canale & Swain 1980, Van Ek 1986, Sijtstra 1991) en het taalgebruiks- en taalleerproces wordt opgevat als een probleemoplossingsproces (cf. Van Dijk & Kintsch 1983, Levelt 1989, Hayes & Flower 1980, McLaughlin 1987).

Het model ligt ten grondslag aan een analyse van enkele veelgebruikte leergangen Nederlands en moderne vreemde talen met als hoofdvraag in welke mate deze leergangen voldoen aan enkele relevant geachte kenmerken van strategie-onderwijs dat transfer binnen en tussen talen faciliteert[2]. Met transfer binnen talen gaat het erom of kennis en vaardigheden aangeleerd in de ene taak ook toegepast worden bij andere (nieuwe) taken. Bij transfer tussen talen gaat het om de vraag in hoeverre kennis en vaardigheden aangeleerd bij taken in de ene taal ook toegepast worden in het verwerven en gebruiken van een andere taal[3]. Op grond van de resultaten van deze leerganganalyses gaan we in de discussie in op de vraag in hoeverre binnen het talenonderwijs aandacht wordt besteed aan de ontwikkeling van strategische competentie.

2. EEN MODEL VOOR TAALGEBRUIK EN TAAL LEREN

Het model (Oostdam & Rijlaarsdam 1995) bestaat uit vier componenten: de input, het werkgeheugen, het lange-termijngeheugen, en de output (figuur 1).

Input

Bij de input is een onderscheid te maken tussen *taaltaken* en *leertaken* (cf. Breen 1987 a&b). Onderscheidend kenmerk van taal- en leertaken is het – al dan niet voorgeschreven – doel van

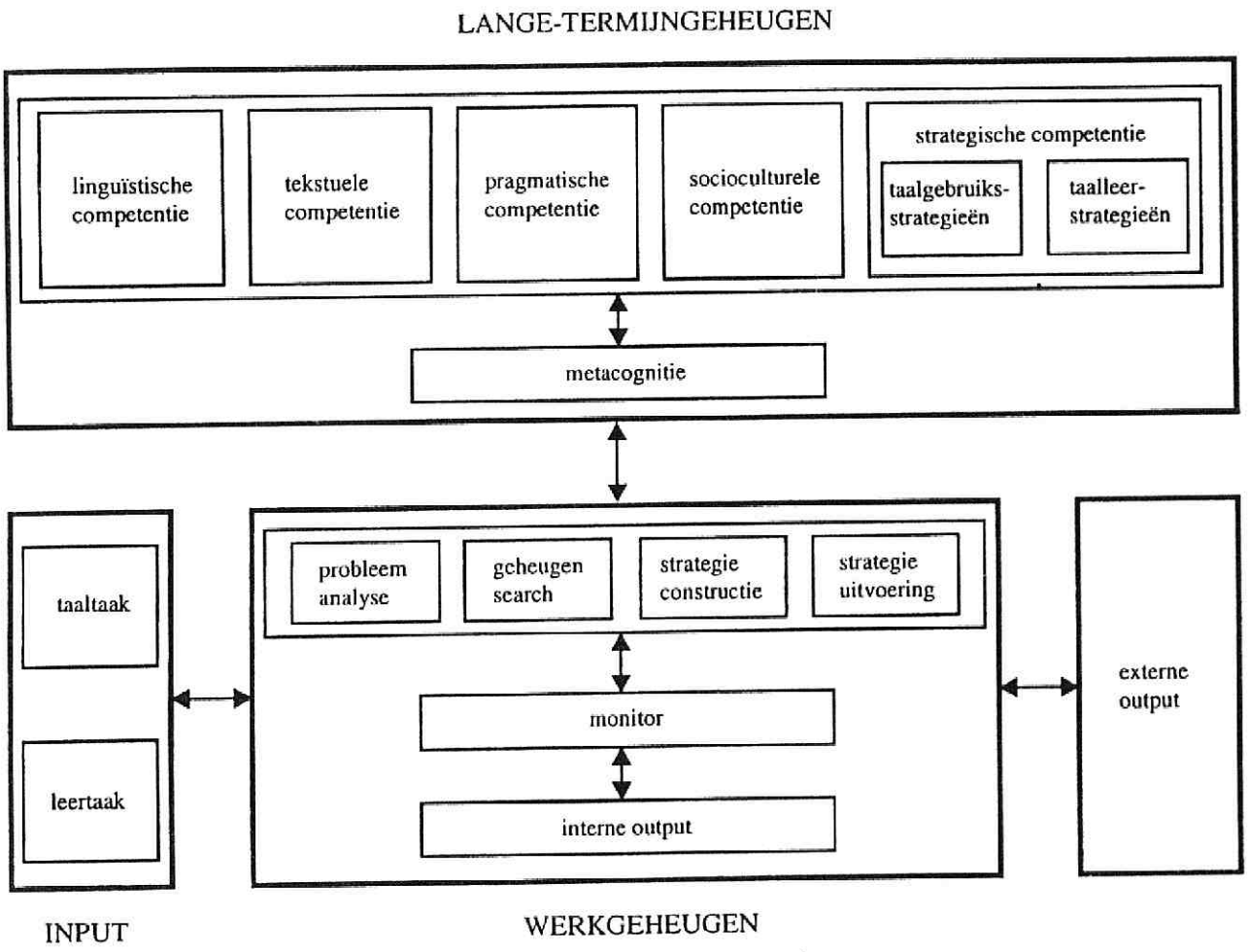


Fig. 1. Model voor taalgebruik en taalleren.

de leerder. Bij een taaltaak is het doel van de leerder effectief communiceren. Bij een leertaak is het doel van de leerder effectief leren. Een herschrijftaak bijvoorbeeld is een taaltaak als het gaat om het verbeteren/corrigeren van eigen of andermans werk binnen een communicatieve context. Het betreft een leertaak als het leren herschrijven het doel is. Dat impliceert niet dat van het uitvoeren van een taaltaak geen leereffect kan uitgaan: er kan sprake zijn van incidenteel leren en daarnaast ziet een goede leerder een taaltaak soms ook als een leertaak.

Lange-termijngeheugen

In het lange termijn geheugen worden in dit model vijf competenties onderscheiden die een functie vervullen in het proces van talig handelen en taalvaardigheidsontwikkeling: de linguïstische, tekstuele, pragmatische, socio-culturele en strategische competentie (cf. Bachman 1990, Canale & Swain 1980, Sijtstra 1991). Het kan daarbij zowel gaan om declaratieve kennis als om kennis in gecompileerde vorm (Van Berkum & De Jong 1991). Declaratieve kennis is in de regel goed expliciet te verwoorden, terwijl dit bij gecompileerde kennis niet perse (meer) het geval hoeft te zijn. Er is sprake van een continuüm: naarmate declaratieve kennis gecompileerder wordt, verandert kennis in vaardigheid.

De *linguïstische competentie* omvat impliciete en expliciete kennis van grammaticale regels, woordbetekenissen, voorschriften voor interpunctie, spelling, uitspraak en articulatie. In hoeverre de linguïstische competentie ontwikkeld is, kan bijvoorbeeld worden afgeleid uit het oordeel van een taalgebruiker over de grammaticale aanvaardbaarheid van een uiting, of uit de score op een woordenschattoets.

Bij de *tekstuele competentie* gaat het om tamelijk gestandaardiseerde conventies ten aanzien van mondelinge of schriftelijke teksten. Essentieel daartoe is het besef van coherentie: bij een tekst gaat het niet om een verzameling losse zinnen, maar om een zodanige ordening dat er een geheel ontstaat. Er is sprake van zekere opbouwpatronen. Besef van coherentie is fundamenteel voor het begrip van verschillende tekstkenmerken, zoals de opmaak van een brief, een kranteartikel en diverse andere tekstschema's. In schematheoretisch georiënteerde literatuur over leesvaardigheid wordt kennis hieromtrent ook wel aangeduid als 'formal schemata' (Carrell 1984).

De *pragmatische competentie* betreft kennis omtrent regels voor taalgebruik en talige interactie, e.g. kennis over hoe taal en woordkeus aangepast worden aan bepaalde situationele contexten en kennis van beleefdheidsprincipes in taalgebruik. De leerder weet bijvoorbeeld dat een telefoongesprek in het Duits wordt afgesloten met 'Auf Wiederhören' – en niet met 'Auf Wiedersehen', of dat Fransen zich aan de telefoon melden met 'Allo' en dat dat niet ongemanierd is.

Bij de *socioculturele competentie* gaat het om kennis over de maatschappelijke context waarbinnen taal functioneert (weten dat bepaalde handelingen al dan niet gepast zijn binnen een bevolkingsgroep). Zo is het in Frankrijk niet gepast iemand feestelijk te verrassen met een boeket witte chrysanten: deze bloemen zijn alleen gepast bij begrafenissen. Wanneer de socioculturele competentie te kort schiet, kan er het nodige mis gaan. Een Nederlander die door Fransen uitgenodigd wordt voor het avondeten en te horen krijgt: 'Ne venez pas trop tard' zou op grond van zijn Nederlandse verwachtingspatroon om ongeveer vijf uur 's middags aanbellen. Grote kans dat hij dan niemand aantreft. De etenstijden in Frankrijk beginnen nu eenmaal later en de mededeling betekent niet vijf uur 's middags maar ongeveer zeven uur 's avonds.

De *strategische competentie* speelt een belangrijke rol bij de receptie en produktie van taal. Op de strategische competentie wordt een beroep gedaan bij het plannen, bewaken, uitvoeren en evalueren van communicatie- en leerprocessen en het eventueel herzien van de plannen.

Bij receptie van taal zullen taalgebruikers bijvoorbeeld, afhankelijk van ondermeer het type tekst, de context, het onderwerp en de eigen voorkennis, meer of minder betekenis kunnen voorspellen (Smith 1988, Westhoff 1981). In geval van een voorspellende leesstrategie heeft de lezer al zekere verwachtingen over een tekst en zal hij tijdens het lezen door middel van 'sampling' controleren of die verwachtingen kloppen. Kloppen de verwachtingen, dan kan de lezer de voorspelstrategie voortzetten; kloppen ze niet, dan kan de lezer beter (tijdelijk) overstappen op een leesstrategie waarbij de perceptie van de tekst zich primair richt op het woord- of zinsniveau: door te lezen wat er staat, wordt dan een representatie van de tekststructuur en tekstinhoud gemaakt.

Bij de produktie van taal kan een taalgebruiker bijvoorbeeld zijn kennis omtrent schrijfprocedures inzetten. Soms kan een schrijver volstaan met een globale strategie om tot tekst te komen (e.g. gegevens verzamelen en ordenen, tekst uitschrijven en nakijken), maar in geval van bijzondere tekstsoorten (een sollicitatiebrief, een beleidstekst) zal de schrijver gebruik kunnen maken van een meer specifiek ingevulde strategie omdat bijvoorbeeld de inhoudselementen en de volgorde min of meer vaststaan.

In navolging van Selinker (1971), Faerch & Kasper (1983 a&b) en Tönshoff (1992) is er binnen de strategische competentie een onderscheid te maken tussen *taalgebruiksstrategieën* en *taalleerstrategieën*: hoe doe ik X versus hoe leer ik X te doen. Gaat het bij taalgebruiksstrategieën om het gebruiken van taal voor communicatieve doeleinden, bij leerstrategieën gaat het om taalverwerving. Voorbeelden van taalleerstrategieën zijn: het vormen van hypotheses over regelmatigheden in een taal en het toetsen van zulke hypotheses (Faerch & Kasper 1983 a&b), het observeren van discussiedeelnemers met het expliciete doel te leren van andermans gedrag, het mentaal visualiseren van woordbetekenissen ter memorisering en het kritisch evalueren van de effectiviteit van gehanteerde strategieën (leren door zelfreflectie).

Metacognitie heeft betrekking op de kennis die iemand heeft over de eigen kennis, over zichzelf en anderen als taalgebruiker en -leerder, evenals kennis over de moeilijkheidsgraad van taken in relatie tot de eigen capaciteiten. Metacognitie ontstaat op basis van metacognitieve ervaringen (Flavell 1976, 1978, 1981, Garner 1987, Elshout- Mohr 1992). Wanneer een taalge-

bruiker weet dat een bepaalde strategie voor hemzelf wel effectief is in de ene situatie, maar niet in een andere, dan is dat metacognitieve kennis. Tussen metacognitieve ervaringen, metacognitie en de cognitieve competenties bestaat een intensieve wisselwerking (zie voorbeelden bij Garner 1987, Elshout-Mohr 1992).

Werkgeheugen

In het model omvat het werkgeheugen de volgende interagerende componenten: probleemanalyse, geheugensearch, strategie-constructie, strategie-uitvoering, de monitor en interne output.

Onder *probleemanalyse* wordt verstaan dat het probleem in de gestelde taal- of leertaak wordt onderkend en ontleed in functionele eenheden. Bij een complex probleem zal de diagnose zich moeten richten op ontleding van het probleem in mogelijke deelproblemen. Tijdens de taakuitvoering kan de probleemdefinitie veranderen. Bijvoorbeeld: een leerling moet een studietekst lezen. Daartoe moet hij de hoofdgedachte van een tekstfragment bepalen. Hij komt echter een centraal woord tegen waarvan hij de betekenis niet kent. Op dat moment ziet hij zich gesteld voor een nieuw probleem en zal hij de betekenis van het woord moeten achterhalen voordat kan worden overgegaan tot een oplossing van de eigenlijke probleemstelling (tekst lezen). Resultaat van een probleemanalyse is een decompositie van het probleem in de gegeven taak en het stellen van een doel (het criterium waar de oplossing aan moet voldoen). Daartoe worden zowel de taak- als de situatiekenmerken (is er bijvoorbeeld een tijdslimiet aan de taakuitvoering) in ogenschouw genomen (cf. Miller, Galanter & Pribram 1960).

De *geheugensearch* is gericht op de mogelijkheden van de cognitieve competenties en de metacognitie om bij te dragen aan de probleemanalyse en de probleemoplossing. Wanneer een taalgebruiker bijvoorbeeld niet beschikt over kennis om woordbetekenissen af te leiden dan stelt dit beperkingen aan zijn leesproces (hij zal zich dan bijvoorbeeld moeten richten op het opzoeken van betekenissen in het woordenboek).

Tijdens de *strategieconstructie* staan verschillende opties open. Na analyse van het probleem, zal een taalgebruiker – gegeven zijn cognitieve condities – een (in het geheugen) beschikbare strategie kunnen toepassen, een beschikbare strategie kunnen modificeren, of een nieuwe strategie kunnen samenstellen.

Als een taalgebruiker bij het lezen van een tekst bijvoorbeeld op een onbekend woord stuit en kennis van het betreffende woord is een noodzakelijke voorwaarde voor een goed begrip van de tekst en de betekenis van het woord is morfologisch afleidbaar en de taalgebruiker beschikt over de relevante morfologische kennis, dan is het mogelijk om een morfologische afleidstrategie toe te passen. Als de morfologische kennis slechts ten dele aanwezig is, moet de taalgebruiker beslissen zijn strategie te modificeren of op een andere strategie over te gaan. Modificatie van de strategie kan eruit bestaan dat de morfologische afleidstrategie gecombineerd wordt met een raadstrategie. Deze raadstrategie kan de morfologische afleidstrategie zelfs verdringen zodra blijkt dat deze niet voldoet. Indien blijkt dat de morfologische afleidstrategie niet uitvoerbaar is dan kan een taalgebruiker beslissen de condities te heranalyseren. Dit heranalyseren kan bijvoorbeeld gebeuren via het toepassen van een (nieuwe) deelstrategie om de cognitieve condities voor het uitvoeren van een morfologische afleidstrategie te optimaliseren (e.g. nagaan of bepaalde kenmerken van de woordbetekenis uit de context zijn af te leiden die de morfologische analyse kunnen ondersteunen).

De *monitor* in het model observeert het probleemoplossingsproces teneinde de effectiviteit ervan te bewaken en bijsturen mogelijk te maken. Dit bewaken en sturen verloopt recursief en interactief. Wanneer de geheugensearch bijvoorbeeld geen goede strategie genereert, kan worden teruggegaan naar de probleemanalyse. Wellicht kan een heranalyse van het probleem plaatsvinden op grond waarvan opnieuw een geheugensearch kan plaatsvinden. Ook wanneer het uitvoeren van een gegenereerde strategie op problemen stuit, kan worden teruggegaan naar eerdere stappen in het probleemoplossingsproces.

Merk op dat begrippen als 'monitoring' en 'zelfregulatie' vaak geplaatst worden onder de paraplu van metacognitie (cf. Elshout-Mohr 1992). Wij maken onderscheid tussen monitoring (proces) en metacognitie (kennis). In het model duidt de monitor op het proces van zelfregula-

tie/monitoring. De configuraties volgens welke de monitor werkt, liggen opgeslagen in het lange termijn geheugen als deel van de metacognitie. Nieuwe configuraties die aldoende worden ontwikkeld, kunnen als metacognitieve kennis worden opgeslagen.

Interne en externe output

Het zal duidelijk zijn dat het geheugen en geheugenvaardigheden een zeer belangrijke rol spelen in taalgebruiks- en taalleerprocessen. Een effectieve *strategie-uitvoering* leidt tot een probleemoplossing. Hierbij is een onderscheid te maken tussen *interne* en *externe output*.

Interne output betreft niet direct observeerbare, mentale output als tussentijds of uiteindelijk resultaat van het probleemoplossingsproces. Interne output als tussentijds resultaat vindt bijvoorbeeld plaats tijdens het lezen en bestaat dan uit het (tijdelijk) opslaan van de boodschap en het zonodig weer ophalen ervan. Tegelijkertijd moeten interpretaties en reacties op informatie uit de tekst worden onthouden. Daarbij moet nieuwe informatie weer kunnen worden opgenomen en de voorlopige representaties worden bijgesteld. Een voorbeeld van interne output als uiteindelijk resultaat is het opslaan van woordbetekenissen in het geheugen als resultaat van een vocabulaire-oefening (interne output die de linguïstische competentie vergroot).

Externe output is direct observeerbaar, bijvoorbeeld de geschreven brief, de gehouden spreekbeurt, de aantekeningen bij een gelezen of beluisterde tekst (ook tijdens het lezen of luisteren).

3. AANDACHT VOOR STRATEGISCHE COMPETENTIE IN LEERGANGEN

Om na te gaan of en in hoeverre er sprake is van taalonderwijs dat zich richt op de strategische competentie en of en in hoeverre dat onderwijs voldoet aan elementaire voorwaarden voor transfer, is een onderzoek verricht naar leerboeken Nederlands (moedertaalonderwijs), Engels en Frans (zie Oostdam en Rijlaarsdam 1993 a&b, 1995). Voor de beschrijving van taken die leerlingen moeten uitvoeren in groep 7 en 8 van de basisschool (Nederlands en Engels) en in de leerjaren 1, 2 en 3 in het voortgezet onderwijs (Nederlands, Engels en Frans) is een instrument ontwikkeld en beproefd dat gebaseerd is op het hiervoor beschreven model. Met behulp van dit instrument werden van elke geselecteerde oefening zes kenmerken vastgesteld.

(1) De cognitieve actie die kennelijk van een leerling werd gevraagd. Het coderingsvoorschrift werd opgesteld op basis van een oefentypologie (Neuner et al. 1981, Van der Voort & Mol 1989).
(2) De modus van communicatie (lezen, schrijven, spreken, luisteren). Gecodeerd werd welke modaliteit kennelijk geoefend werd. Als de modaliteit niet expliciet werd aangegeven en de docent dus vrijheid van handelen had om een taak mondeling dan wel schriftelijk te doen, vond geen scoring op deze categorie plaats.
(3) Het domein van taalkennis. Het scoringsinstrument bevatte dezelfde onderscheidingen als de in het model onderscheiden competenties.
(4) De kennelijke gerichtheid op een taalgebruiks- en/of taalleerstrategie. Gescoord werd of de oefening gericht was op het aanleren van een taalgebruiksstrategie, een taalleerstrategie, op beide of geen van beide. Het onderscheid tussen taalgebruiks- en taalleerstrategie kwam in de scoring overeen met de in paragraaf 2 gemaakte onderscheiding. Werd er een strategie aangeboden om een taalverkeersprobleem op te lossen (taalgebruiksstrategie) of werd er een strategie aangeboden om een leerprobleem op te lossen (leerstrategie)?
(5) De aanwezigheid van een formulering die het onderwijsdoel aangaf en die een aanwijzing bevatte over de toepassingscontext. De formulering moet ten minste het onderwijsdoel expliciteren wil de oefening een score op deze categorie krijgen. Er is geen onderscheid gemaakt tussen expliciteren en legitimeren, hoewel in het laatste geval de transfercontext duidelijker is dan in het eerste geval.
(6) De aanwezigheid van interne controle-opdrachten, die onderwijs in controlestrategieën of metacognitie kunnen indiceren. Aanwezigheid van dit kenmerk duidt erop dat de leergang de leerling steun biedt bij het uitvoeren van de taak.

In deze bijdrage beperken we de rapportage tot vier samenhangende onderzoeksvragen. De eerste drie vragen betreffen de mate waarin er sprake is van strategie-onderwijs binnen de onderscheiden vakken. Als deze vragen positief worden beantwoord en er dus voldoende aanwijzingen zijn voor het bestaan van dergelijk onderwijs, stellen we de vierde vraag, in hoeverre dit onderwijs zich richt op transfer.

1. *Leereffect*: is de taak gericht op een taalgebruiksstrategie, een taalleerstrategie, zowel de een als de ander of geen van beide?

Deze vraag betreft de inhoud van het taalonderwijs: is er aandacht voor de strategische competentie? (Zie par. 2.) De vraag kan beantwoord worden doordat van elke taak is vastgesteld welke leeractiviteit de taak uitlokte (analysecategorie 1, zie hierboven) en, onafhankelijk daarvan, of de taakinhoud kennelijk een leerstrategie en/of een taalgebruiksstrategie beoogde (analysecategorie 4, zie hierboven). Alle leeractiviteiten of type oefeningen hebben we ondergebracht in twee categorieën: leer- of taaltaken (zie tabel 1). Sommige oefeningen, zoals vocabulaireoefeningen, doen een beroep op cognitieve activiteiten (inprenten, sorteren, thematiseren) die van belang zijn bij het leren verwerven van de vreemde taal; we categoriseren zo'n oefening dan als potentiële leerstrategie. Potentieel, omdat andere kenmerken van de oefening de potentie omzetten naar een actuele leerstrategie. Als de leerling niet alleen de opdracht krijgt 'leer deze woordjes' maar ook een strategie aangereikt krijgt hoe deze woordjes te leren zijn, dan spreken we van een leerstrategie. Een ander voorbeeld is het gebruik van het woordenboek in een taak. Ook die taak heeft de potentie van een leerstrategie: een leerling leert dat via het woordenboek bepaalde verwervingsproblemen op te lossen zijn. Het hangt af van de specifieke taak in het leerboek of leerlingen inderdaad een leerstrategie aangeboden krijgen of dat zij (louter) een woordenboekoefening uitvoeren.

2. *Leeractiviteiten*: stimuleert de taak 'monitoring'-activiteiten of zelfregulatie-activiteiten?

Deze vraag gaat in op het probleemoplossingsaspect van het in paragraaf 2 gepresenteerde model: leren leerlingen hun taakuitvoering te beschouwen, te controleren, te observeren? Leren leerlingen de taakuitvoering te besturen? De vraag kan beantwoord worden doordat van elke oefening is bepaald of de oefening aanwijzingen bevatte voor de leerling waarmee hij de taakuitvoering kon beoordelen, zowel tijdens als na het uitvoeringsproces (zie categorie 6, hierboven).

Van elke oefening is aangegeven of er sprake is van enige sturing door de leerboekauteurs: planningsaanwijzingen, decompositie-aanwijzingen of -opdrachten, tussentijdse controles, eindcontroles, reflectieopdrachten. Daarbij is geen onderscheid gemaakt tussen monitoring en zelfregulatie. Onderwijs in dergelijke activiteiten wordt door ons beschouwd als een belangrijke bijdrage aan het vervullen van transfervoorwaarden, doordat leerlingen het uitvoeringsproces (zie het werkgeheugen in figuur 1) gemodelleerd krijgen en gelegenheid krijgen om een metacognitieve ervaring op te doen.

3. *Leerdoelexplicitering*: bevat de taak een expliciete vermelding van het leerdoel of de transfercontext van de getrainde strategie?

Deze vraag heeft net als vraag 2 betrekking op transfervoorwaarden: in de inleiding is reeds naar voren gebracht dat de kans op transfer naar andere taken verhoogd kan worden door (ten minste) expliciete verwijzing naar transfercontexten. We kunnen de onderzoeksvraag beantwoorden door de scores op de categorie 5 van hierboven in beschouwing te nemen. Van elke taak is dan ook vastgesteld of de taak een formulering bevat die het onderwijsdoel aangeeft en/of een aanwijzing bevat over de toepassingscontext.

4. *Transfer tussen schoolvakken*: als zowel in het moedertaalonderwijs als het vreemde-talenonderwijs sprake is van onderwijs gericht op de strategische competentie, en dat onderwijs voldoet aan de twee onderscheiden voorwaarden (zie de vragen 2 en 3) voor transfer, is er dan kans op transfer tussen de talen?

Om deze vraag te kunnen beantwoorden is van elke taak vastgesteld welke van de competenties (pragmatisch, tekstueel, linguïstisch) en welke van de modaliteiten (lezen, luisteren, spreken en schrijven) in het geding was (categoriëen 2 en 3), zie hierboven. De aanname hierbij is dat taken van verschillende vakken die gericht zijn op de strategische competentie èn die qua competentie en/of modaliteit overeenkomen, kans op transfer tussen vakken indiceren.

Onderzoeksmateriaal

Onderzocht is lesmateriaal voor de twee hoogste groepen van het basisonderwijs en de eerste drie leerjaren van het voortgezet onderwijs. Voor het basisonderwijs zijn twee leergangen voor Nederlands en twee leergangen voor Engels onderzocht[4]. Voor het voortgezet onderwijs zijn drie leergangen Nederlands, drie leergangen Engels en twee leergangen Frans onderzocht[5]. De leergangen zijn geselecteerd op basis van destijds (1991) beschikbare gegevens over het marktaandeel (Edelenbos, Suhre & Zuidema 1989, Kuhlemeier 1989, Kuhlemeier & Van Werkhoven 1991). We hebben ons beperkt tot het analyseren van leergangen voor twee moderne vreemde talen. Gekozen is voor Engels om een vergelijking tussen leerstof en didactiek in basis- en voortgezet onderwijs te kunnen maken en voor Frans om de vergelijking tussen een taal waarmee leerlingen in het voortgezet onderwijs al wel (Engels) en niet (Frans) vertrouwd zijn.

Van elke leergang zijn alle oefeningen uit de eerste en laatste twee hoofdstukken opgenomen in de steekproef van te coderen oefeningen. Eenheid van codering was de eenheid waarin de leergang de oefeningen aanbood[6]. Alleen de kernoefeningen werden in de steekproef opgenomen: differentiatie- en herhalingsoefeningen werden van codering uitgesloten evenals toetsen. De codering is ruimhartig toegepast: bij twijfel werd het te coderen verschijnsel als aanwezig gecodeerd.

Een steekproef van oefeningen werd door twee codeurs gecodeerd. De beoordelaarsbetrouwbaarheid uitgedrukt als het percentage overeenkomstig gescoorde opdrachten was ruim voldoende[7].

Resultaten

Tabel 1 bevat de percentages van het *potentieel aan strategieën*. Van elke oefening is vastgesteld welke leeractiviteit kennelijk beoogd werd. De codering was afgeleid van een indeling van taaloefeningen (cf. Neuner et al. 1981, Van der Voort & Mol 1989). Ons onderscheid in taal- en leertaken voegden wij toe. Een zogenaamde recombinatieoefening waarin leerlingen reeds verworven taalfragmenten combineren op het niveau van de zin of hoger, behoort bijvoorbeeld tot de taalleerstrategieën. Want in principe verricht een leerling hier een activiteit (recombineren) die hem later, bij het zelfstandig dóórleren, van nut kan zijn. Immers, als men zelfstandig met nieuwe taalelementen recombineert, dan oefent men zich in het memoriseren van een constructie in verschillende taalomgevingen. De beschikbaarheid en de bruikbaarheid van het verworven taalelement nemen daarmee toe.

Tabel 1 laat zien dat een beperkt aantal cognitieve activiteiten volstaat om alle oefeningen in leergangen in te delen. De totale percentages *potentiële* taal- en leertaken laten zien dat er in het basisonderwijs zowel voor Nederlands als Engels min of meer gelijke aandacht is voor taal- en leertaken. In het voortgezet onderwijs neemt het aandeel van leertaken toe. Voor Nederlands vindt er een omslag plaats: de verhouding tussen leer- en taaltaken verandert grofweg van 2:3 in 3:2 (37.4:58.8 en 57.7:36.7). Voor Engels neemt de nadruk op leertaken in het voortgezet onderwijs ook toe. Tussen Engels en Frans zijn er in het voortgezet onderwijs op het niveau van leer- en taaltaken geen opvallende verschillen. De verschillen tussen Engels in het basisonderwijs en Engels in het voortgezet onderwijs wijzen waarschijnlijk op een verschil in didactische uitgangspunten.

Het is nu de vraag of de activiteiten die leerlingen dienen uit te voeren, en die wij bestempelden als potentieel leer- of taalstrategisch onderwijs, ook werkelijk als zodanig worden benut. Want de eerste onderzoeksvraag betreft de mate waarin leerlingen daadwerkelijk een taalgebruiks- en/of leerstrategie aangeboden krijgen. Daarvoor kijken we in tabel 2.

De resultaten in tabel 2 laten zien dat leerlingen alleen in leergangen Nederlands taalge-

Tabel 1. Het potentieel aan strategieën in het taalonderwijs. Percentages per schooltype in leergangen Nederlands (Ned.), Engels als vreemde taal en Frans als vreemde taal voor basisonderwijs (leerjaar 7 en 8) en voortgezet onderwijs (leerjaar 1, 2 en 3).

	Basisschool		Voortgezet onderwijs		
	Ned.	Engels	Ned.	Engels	Frans
Leertaken	37.4	51.3	57.7	70.6	66.2
Woordverwerven	7.2	15.5	6.5	14.8	8
Woordenboek gebruiken	2.9	1	3.8	–	–
Parafraseren (woord)	.5	–	.6	2.8	5.1
Reproduceren	10.0	11.4	7.6	16.6	10.8
Recombineren	12.8	8.0	37.4	28.0	26.6
Waarnemen (horen, onderscheiden)	1.2	3.0	–	4.2	3.0
Afleiden woordbetekenissen	2.5	.5	1.4	1.5	–
Uitspraak	0.3	11.9	.4	2.7	12.7
Taaltaak	58.8	47.5	36.7	29.4	35.9
Herkennen taalfunctie	.3	3.7	.3	2.0	1.4
Structureren tekst	5.8	1.0	4.7	.2	–
Anticiperen op onderwerp	3.6	.5	.8	.1	–
Voorspellen (inhoud)	1.3	–	–	–	–
Waarnemen (globaal)	1.3	–	.3	.6	.3
Begrijpen (tekst)	8.5	18.1	16.1	17.3	20.4
Kritisch reflecteren	6.9	.5	4.9	1.2	–
Produceren van tekst	2.3	2.9	3.0	1.3	2.7
Genereren van inhoud	13.0	4.2	4.1	.7	–
Reconstrueren	–	1.7	.1	1.6	5.3
Communiceren	15.8	14.9	2.4	4.4	5.8
Rest	4.9	1.8	5.6	.1	1.1
Aantal oefeningen	366	304	1254	1096	371
Aantal leergangen	2	2	3	3	2

bruiksstrategieën krijgen onderwezen (in het basisonderwijs bevat 27% van de oefeningen een taalgebruiksstrategie en in het voortgezet onderwijs 20%). Opvallend is de grotere aandacht in het basisonderwijs. Aandacht voor leerstrategieën is vrijwel alleen in leergangen Nederlands aangetroffen, en wel in een zeer beperkte mate: 2.3% van de oefeningen in het basisonderwijs en 1.4% in het voortgezet onderwijs.

We zijn ook nagegaan welke strategieën leerlingen kunnen verwerven in de leergangen Nederlands door de gegevens van tabel 1 en 2 te combineren: welke van de cognitieve activiteiten (tabel 1) wordt als een taalgebruiks- of leerstrategie aangeboden?. Opvallend resultaat was het verschil tussen het basisonderwijs en het voortgezet onderwijs. Leerlingen in het basisonderwijs krijgen vooral taalgebruiksstrategieën aangeboden met betrekking tot het tekstuele domein zoals anticiperen, tekststructuren bepalen en het genereren van ideeën (bij schrijven). In het voortgezet onderwijs blijkt een aanzienlijk deel van de taalgebruiksstrategieën te liggen op het terrein van de linguïstische kennis: het oplossen van zinsbouwproblemen. Ook deze verschillen duiden op een verschil in didactische optiek tussen basis- en voortgezet onderwijs.

Dat van de potentiële taalgebruiks- en leerstrategieën (zie tabel 1) in leergangen voor het vreemde-talenonderwijs er maar zo weinig werkelijk benut worden (maximaal 5.1% bij Engels in het basisonderwijs, zie tabel 2), betekent dat de kans op transfer tussen onderwijs Nederlands

Tabel 2. Relatief aantal oefeningen (percentages) met taalgebruiksstrategie, taalleerstrategie, transfercontext en monitoringactiviteiten in leergangen Nederlands (Ned.), Engels als vreemde taal en Frans als vreemde taal voor basisonderwijs (leerjaar 7 en 8) en voortgezet onderwijs (leerjaar 1, 2 en 3).

	Basisonderwijs		Voortgezet onderwijs		
	Ned.	Engels	Ned.	Engels	Frans
taalgebruiksstrategie (1)	26.5	5.1	19.8	.8	–
leerstrategie (2)	2.3	.3	1.4	–	–
transfercontext (3)	2.8	–	1.6	–	.2
monitoring (4)	16.9	.3	7.7	.1	
Combinaties:					
1 + 2 + 3 + 4	.3	–	.3	–	–
1 + 3 + 4	1.3	–	.6	–	–
1 + 3	.8	–	.5	–	–
1 + 4	7.8	–	3.7	.1	–
2 + 3 + 4	–	–	–	–	–
2 + 3	–	–	–	–	–
2 + 4	–	–	.1	–	–
1 + 2	1.5	.3	.3	–	–
3 + 4	.3	–	.1	–	–
Aantal oefeningen	366	304	1254	1096	371
Aantal leergangen	2	2	3	3	2

en vreemde-talenonderwijs met het huidige lesmateriaal zeer gering is. Onderzoeksvraag 4 is daarmee beantwoord.

De tweede en derde onderzoeksvraag betreffen kenmerken van transferbevorderend talenonderwijs. Het gaat daarbij om de mate waarin 'monitoring'-activiteiten of zelfregulatieactiviteiten worden gestimuleerd en om de expliciete vermelding van het leerdoel of de tranfercontext.

Uit tabel 2 zijn de resultaten af te lezen met betrekking tot deze twee didactische kenmerken. De proportie taken met een expliciet verwoord leerdoel of een toepassingscontext is gering (.028). De proportie taken waarin leerlingen op de een of andere wijze steun krijgen bij de taakuitvoering is weliswaar groter (.169 in het basisonderwijs en .077 in het voortgezet onderwijs), maar toch gering. Als er dus al strategiegericht onderwijs is qua inhoud, dan faciliteert de aanbiedingsvorm transfer naar andere taal- en leertaken waarschijnlijk niet.

De leeractiviteiten die begeleid worden via de leergang in het basisonderwijs verschillen van die in het voortgezet onderwijs: in het basisonderwijs betreft het de activiteiten communiceren en kritisch reflecteren, in het voortgezet onderwijs betreft het vooral recombineren[8]. In het basisonderwijs hebben begeleidende teksten voor taakuitvoering vooral betrekking op het tekstuele domein, in het voortgezet onderwijs vooral op het linguïstische domein. Wederom lijken deze gegevens een verschil in didactische opvatting te reflecteren: het voortgezet onderwijs richt zich in kwantitatieve zin meer op kleinere eenheden (zinnen) en eenvoudige cognitieve activiteiten (recombineren), terwijl het basisonderwijs aandacht schenkt aan grotere gehelen (teksten) en complexere cognitieve activiteiten (reflecteren).

Als we de beoogde leereffecten en didactische kenmerken combineren (zie tabel 2, tweede helft) dan kunnen we ten aanzien van de vraag naar transfer drie conclusies trekken.

1. In het basisonderwijs wordt in de leergangen Nederlands in een flink aantal gevallen waarin leerlingen een taalgebruiksstrategie wordt aangeboden ook aandacht besteed aan zelfregulatie: 9.4% (de combinaties 1+3+4; 1+2+3+4; 1+4) van het totaal aantal oefeningen of ongeveer in een derde van de gevallen (9.4/26.5) waarin een taalgebruiksstrategie wordt aangeboden. In het voortgezet onderwijs komt dat veel minder vaak voor: 4.6% (0.3+0.6+3.7) van het

totaal of een vierde (4.6/19.8) van de gevallen waarin een taalgebruiksstrategie wordt aangeboden. Het aantal gevallen waarin leerlingen een leerstrategie wordt aangeboden met ondersteuning (monitoring) is verwaarloosbaar (alle combinaties met 2 en 4 erin).

2. In het basisonderwijs en in het voortgezet onderwijs krijgen leerlingen maar heel zelden taalgebruiks- of leerstrategieën aangeboden met een toepassingscontext (combinaties met 3). In het basisonderwijs betreft dat nog geen tiende van de gevallen waarin een strategie wordt aangeboden, in het voortgezet onderwijs nog veel minder.
3. Het aantal opdrachten waarin leerlingen zowel een taalgebruiks- als een leerstrategie verwerven (de combinaties met 1 en 2), is verwaarloosbaar klein. Dit zijn oefeningen waarin leerlingen dus niet alleen bijvoorbeeld een lees- of schrijfstrategie verwerven, maar ook leren een dergelijke strategie te verwerven.

4. BESPREKING

Ten aanzien van het vak Nederlands dringen twee conclusies zich op. (1) Leerboeken Nederlands bieden leerlingen strategieën aan om communicatieproblemen op te lossen, maar nauwelijks leerstrategieën. (2) Het aanbod in het basisonderwijs voor het vak Nederlands is zowel kwantitatief (aandeel van strategisch gericht onderwijs) als kwalitatief (didactische kenmerken) beter. Als we als kwalitatieve eis stellen dat lesmateriaal gericht op het aanleren van een strategie ten minste één didactisch kenmerk moet vertonen (transfercontext of metacognitieve controle) dan voldoet zo'n 10% van de oefeningen in het basisonderwijs en zo'n 5% van de oefeningen in het voortgezet onderwijs daaraan. Opvallend daarbij is dat het accent verschuift: in het basisonderwijs ligt de nadruk op strategieën met betrekking tot het tekstuele domein, in het voortgezet onderwijs op strategieën met betrekking tot het linguïstische domein.

De toestand van het moderne vreemde-talenonderwijs is inhoudelijk en didactisch slechter: leergangen voor de moderne vreemde talen zijn in het geheel niet toegerust voor strategieonderwijs omdat zij (1) inhoudelijk niet gericht zijn op strategieën en (2) didactische kenmerken ontberen die transfer zouden kunnen bevorderen. 'Blind training' is hier het hoofdkenmerk van de didactiek. Leeractiviteiten die potentieel leerstrategisch zijn zoals recombineren (zie tabel 1) worden niet als zodanig benut.

Bij deze conclusies is ten minste één belangrijke kanttekening te plaatsen. De resultaten die hier getoond worden, zijn slechts een indicatie van de mate waarin het onderwijs zich richt op strategieën. Een docent heeft natuurlijk de mogelijkheid om oefeningen in leerboeken te voorzien van begeleidende tekst: hij kan leerlingen een aanpak van de oefening suggereren, hij kan transfercontexten toevoegen, hij kan leerlingen leren hun proces en produkt tussentijds te evalueren. Constructie-onderzoek dat momenteel verricht wordt, is erop gericht een heuristiek op te stellen die het mogelijk maakt oefeningen te lokaliseren die zich lenen tot strategische toevoegingen door aanvulling of omzetting. Het blijkt mogelijk om dergelijke potentiële oefeningen te herkennen. Het blijkt ook mogelijk om met een beperkt arsenaal aan middelen dergelijke oefeningen om te bouwen tot oefeningen waarin leerlingen taalgebruiks- en/of leerstrategieën kunnen verwerven.

Het is waarschijnlijk niet gemakkelijk bestaande oefeningen in leergangen zo vorm te geven, dat er sprake is van strategie-onderwijs (zie voor voorbeelden Oostdam & Rijlaarsdam 1993 a&b). Meer valt wellicht te verwachten van docentinterventies, zoals hierboven beschreven. Een gegeven uit dit onderzoek ondersteunt die verwachting: in Taalkabaal is veel meer strategie-onderwijs gesignaleerd dan in andere leergangen. Taalkabaal bestaat nu juist niet alleen uit een leerlingboek, maar veel lessen zijn vormgegeven via docentbladen, die veel mogelijkheden scheppen voor didactische aanwijzingen. Maar de aanwezigheid van docentinterventies maakt niet het enige verschil uit in de mate waarin leergangen aandacht besteden aan strategie-onderwijs. Er zijn ook grote verschillen aangetroffen in de mate waarin het leerlingenmateriaal in de leergangen aandacht besteedt aan strategie-onderwijs. (Een weergave van de resultaten per leergang is terug te vinden in Oostdam & Rijlaarsdam 1993 a&b.)

Een tweede kanttekening betreft de generatie leergangen. De leergangen die in dit onderzoek zijn opgenomen zijn op één na allemaal pre-basisvormingsleergangen. In dit onderzoek hebben we slechts één leergang Nederlands voor de basisvorming (Nieuw Nederlands) geanalyseerd. Als we die scores vergelijken met de scores van de twee pre-basisvorming leergangen Nederlands voor de eerste drie leerjaren van het voortgezet onderwijs dan blijkt de nieuwste leergang inderdaad hoger te scoren op de aandacht voor taalgebruiks- en taalleerstrategieën.

Al met al zijn de resultaten van onze analyses niet erg bemoedigend. Hoewel we aannemen dat de huidige generatie basisvormingsleergangen gemiddeld iets beter zullen scoren op kenmerken van strategisch en op transfer gericht onderwijs, verwachten we pas bij de tweede generatie basisvormingsleergangen een flinke doorbraak in deze.

NOTEN

1. In dit artikel gebruiken wij de termen taaltaak en leertaak als synoniemen voor taalgebruikstaak en taalleertaak.
2. Dit onderzoek is uitgevoerd in het kader van een strategisch onderzoekprogramma naar transfer binnen het talenonderwijs. Dit strategisch onderzoekprogramma wordt uitgevoerd door de onderzoeksgroep Taalonderwijs van het SCO/Kohnstamm Instituut van de Universiteit van Amsterdam met subsidie van het Instituut voor Onderzoek van het Onderwijs (SVO). Daarbij wordt samengewerkt met de vakgroep Taalkunde van de Vrije Universiteit en de universitaire lerarenopleidingen van de Universiteit van Amsterdam (ILO), de Vrije Universiteit (IDO) en de Universiteit Utrecht (IVLOS). Hoofddoel is te onderzoeken welke instructie-maatregelen transfer binnen en tussen talen kunnen bevorderen. Het gaat daarbij in de eerste plaats om talenonderwijs aan van huis uit Nederlandstalige leerlingen.
3. Een beknopte, eerste rapportage van de verschillende (deel)onderzoeken die plaatsvinden in het kader van het strategisch onderzoekprogramma is terug te vinden in: R. Oostdam (Ed.): Transfer in het talenonderwijs, *Levende Talen* 485 (themanummer), december 1993.
4. Voor de basisschool zijn de twee meest gebruikte leergangen voor Nederlands "Taal actief" en "Taal totaal". Deze twee leergangen hebben een marktaandeel van rond de zestig procent (Kuhlemeier 1989). Voor de analyse hebben we één van deze leergangen als representant gekozen. Daarnaast hebben we een basisschoolleergang gekozen waarin volgens didactische literatuur nieuwe didactische inzichten zouden zijn verwerkt (Taalkabaal). Voor Engels (dat verplicht is in groep zeven en acht) zijn twee veelgebruikte leergangen "Engels Basisonderwijs" en "Real English" (Edelenbos, Suhre & Zuidema 1989).
5. In de onderbouw van het voortgezet onderwijs worden "Taalcirkel" (op lbo/mavo) en "Functioneel Nederlands" (op havo/vwo) veel gebruikt (Kuhlemeier 1989; Oostdam 1991). Naast deze twee leergangen hebben we ook de meer recent verschenen leergang "Nieuw Nederlands" in de analyses betrokken. Voor Engels betreft het "Mainstream", "Unicom" en "In Focus" (Kuhlemeier & Van Werkhoven 1991), voor Frans gaat het om "On y va tous" en "Bienvenue en France" (Kuhlemeier & Van Werkhoven 1991).
6. Er is zoveel mogelijk afgegaan op de opdrachtindeling die in de leergang wordt gehanteerd; wat in een leergang als één opdracht of één taalles (zoals in leergangen voor de Basisschool) wordt gepresenteerd is ook als zodanig gescoord. Wel zijn binnen één opdrachtkader (leskader) duidelijk van elkaar te onderscheiden opdrachten als afzonderlijke opdrachten gescoord. Een uitzondering is gemaakt voor de leestekst met vragen: hoewel er vaak sprake is van verscheidene opdrachten (vragen) per tekst, is de leestekst met vragen als één opdracht te beschouwen. Het aantal onderzochte oefeningen per leergang bedraagt: Taal totaal 165; Taalkabaal 201; Functioneel Nederlands 497; Taalcirkel 394; Nieuw Nederlands 363; Engels Basisonderwijs 108; Real English 196; In Focus 279; Unicom 452; Mainstream 365; On y va tous 192; Bienvenue en France 179.
7. De beoordelaarsovereenstemming (uitgedrukt in het percentage overeenstemming) per gescoord kenmerk bedroeg .93 voor taalgebruiksstrategie, .98 voor leerstrategie, .94 voor leerdoel en .98 voor monitoring of zelfregulatie-activiteiten. Daarbij moet wel in ogenschouw worden genomen dat deze percentages overeenstemming geflatteerd zijn door een fikse oververtegenwoordiging van nulscores (het niet voorkomen van een genoemd kenmerk).
8. Recombinatie-oefeningen hebben als doel leerlingen te trainen in het memoriseren van woorden en grammaticale structuren. In recombinatie-oefeningen gaat het erom leerlingen taal te laten gebruiken die in lichte mate afwijkt van het geleerde. Leerlingen moeten bijvoorbeeld woorden of zinsdelen vervangen, zinnen uitbreiden of ombouwen, werkwoordstijden veranderen.

LITERATUUR

Bachmann, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Berkum, J.J.A. van & Jong, T. de (1991). Instructional environments for simulations. In: T. de Jong (Ed.). *Special issue on computer simulations in a instructional context*. Eindhoven/Amsterdam: The SAFE-project, 305-351

Bialystok, E. (1990). *Communication strategies. A psychological analysis of second-language use*. Oxford: Basil Blackwell.

Bimmel, P. (1992). Leren leren in de basisvorming moderne-vreemde-talen. *Levende Talen 472*, 297-304.

Bimmel, P. (1993). Lernstrategien im Deutschunterricht. *Fremdsprache Deutsch 8*, 4-11.

Bimmel, P. & Westhoff, G. (1993). Leesstrategieën: training in het moedertaalonderwijs, transfer naar het vreemde-talenonderwijs. *Levende Talen 485*, 577-581.

Breen, M. (1987a). Contemporary paradigms in syllabus design. Part 1. *Language Teaching* (April).

Breen, M. (1987b). Contemporary paradigms in syllabus design. Part 2. *Language Teaching*, (July).

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics 1*, 1-47.

Carrell, P. (1984). Evidence of a formal schema in second language comprehension. *Language Learning 34* (2), 87-112.

Commissie Herziening Eindtermen (1990). *Advies kerndoelen voor de basisvorming in basisonderwijs en voortgezet onderwijs*. Zoetermeer: O&W/DOP.

Dijk, T. van & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Edelenbos, P., Suhre, C.J.M. & Zuidema, B. (1989). *Methoden voor Engels nader bekeken; Eindrapportage V-EIBO-project*. Groningen: RION.

Van Ek, J.A. (1986). *Objectives for foreign language learning. Volume I: Scope*. Strasbourg: Council of Europe.

Elshout-Mohr, M. (1992). Metacognitie van lerenden in onderwijsleerprocessen. *Tijdschrift voor Onderwijsresearch 17*, 273-289.

Faerch, C. & Kasper, G. (1980). Processes in foreign language learning and communication. *Interlanguage Studies Bulletin 5*, 47-118.

Faerch, C. & Kasper, G. (1983a). *Strategies in interlanguage communication*. London: Longman.

Faerch, C. & Kasper, G. (1983b). Procedural knowledge as a component of foreign language learner's communicative competence. In H. Bolte, & W. Herrlitz (eds.). *Kommunikation im (Sprach-)Unterricht*. Utrecht: Rijksuniversiteit.

Faerch, C. & Kasper, G. (1986). Strategic Competence in Foreign Language Teaching. In G. Kasper. *Learning, Teaching and Communication in the Foreign Language Classroom*. Århus, 179-193.

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 Associates.

Flavell, J.H. (1978). Metacognitive development. In J.M. Scandura & C.J. Brainerd (Eds.). *Structural/process models of complex human behavior*. (pp. 213-245). Alphen aan den Rijn: Sijthoff & Noordhoff.

Flavell, J.H. (1981). Cognitive monitoring. In W.P. Dickson (Ed.). *Children's oral communication skills*. (pp. 35-60). New York etc.: Academic Press.

Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, N.J.: Ablex.

Hayes, J.R. & Flower, L.S. (1980). Identifying the organization of writing processes. In L.W. Gregg & E.R. Steinberg (Eds.). *Cognitive processes in writing*. (pp. 3-30). Hillsdale N.J.: Lawrence Erlbaum Ass.

Hout-Wolters, B. van (1992). *Cognitieve strategieën*. Inaugurale rede. Groningen: Wolters-Noordhoff.

Kuhlemeier, H. (1989). *Inventarisatie van leergangen in mavo, havo en vwo*. Arnhem: Cito.

Kuhlemeier, H. & Werkhoven, R. van (1991). Inventarisatie van leergangen talen in het voortgezet onderwijs. *Levende Talen 457*, 34-37.

Larkin, J.H. (1989). What kind of knowledge transfers? In L.B. Resnick (Ed.). *Knowing, learning and instruction*. (pp. 283-305). Hillsdale, N.J.: Lawrence Erlbaum.

Levelt, W. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

McLaughlin, B. (1987). *Theories of second language learning*. London/New York: Edward Arnold.

Miller, G., Galanter, E. & Pribram, K. (1960). *Plans and the structure of behavior*. New York.

Neuner, G., Krüger, M. & Grewer, U. (1981). *Übungstypologie zum kommunikativen Deutschunterricht*. Berlin: Langenscheidt.

O'Malley, J.M. & Chamot, A.U. (1990). *Learning strategies in second language acquisition*. Cambridge, Cambridge: University Press.

Oostdam, R.J. (1991). *Argumentatie in de peiling; Een aanbod- en prestatiepeiling van argumentatievaardigheden in het voortgezet onderwijs.* Amsterdam: SCO/Universiteit van Amsterdam.
Oostdam, R. (Ed.) (1993). Transfer in het talenonderwijs. *Levende Talen, 485* (themanummer transfer).
Oostdam, R. & Rijlaarsdam, G. (1993a). Strategisch georinteerd moedertaalonderwijs. *Spiegel 11* (1), 9-30.
Oostdam, R. & Rijlaarsdam, G. (1993b). Aandacht voor transfer in leergangen Nederlands en moderne vreemde talen. *Levende Talen 485*, 570-575.
Oostdam, R. & Rijlaarsdam, G. (1995, forthcoming). *Towards Strategic Language Education.* Amsterdam: Amsterdam University Press
Oxford, R.L. (1990). *Language learning strategies; What every teacher should know.* Boston: Heinle & Heinle.
Parreren, C.F. van (1974). Het functioneren van leerresultaten. In C.F. van Parreren & J. Peeck (eds.). *Informatie over leren en onderwijzen.* (pp. 114-130). Groningen: Tjeenk Willink.
Parreren, C.F. van (1981). Onderwijs als vorm van oriëntering van de leerlingen. *Handboek onderwijspraktijk.* Meppel: van Loghem-Slaterus.
Poulisse, N. (1986). 'Niet leren, maar stimuleren!'. *Levende Talen 410*, 229-231.
Poulisse, N., Bongaerts, Th. & Kellerman, E. (1987). Onderzoek naar compensatiestrategieën: enige implicaties voor het vreemde-talenonderwijs. *Levende Talen 426*, 698-704.
Salomon, G. & Globerson, T. (1987). The role of mindfulness in learning and transfer. *International Journal of Educational Research 11*, 623-637.
Salomon, G. & Perkins, D.N. (1989). Rocky roads to transfer. Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist 24*, 113-142.
Schouten-van Parreren, C. (1985). *Woorden leren in het vreemde-talenonderwijs.* Apeldoorn: van Walraven.
Selinker, L. (1971). The psychologically relevant data of second language learning. In P. Pimsleur & T. Quinn, (Ed.). *The psychology of second language learning.* Cambridge: At the University Press.
Sijtstra, J. (1991). *Doel en inhoud van taalonderwijs; De ontwikkeling van een model voor domeinbeschrijvingen van taalonderwijs.* Arnhem, CITO.
Simons, P.R.J. (1988). Leren doen ze zelf. In L.F.W. de Klerk, P.R.J. Simons & J. Zuylen (red.). *Huiswerkbeleid.* (pp. 70-105). Heerlen: Mesoconsult.
Simons, P.R.J. (1990). *Transfervermogen.* Inaugurale rede. Nijmegen: Katholieke Universiteit Nijmegen.
Smith, F. (1988). *Understanding reading. A psycholinguistic analysis of reading and learning to read.* Hillsdale, NJ: Lawrence Erlbaum.
Tönshoff, W. (1992). *Kognitivierende Verfahren im Fremdsprachenunterricht.* Hamburg: Verlag Dr. Kovács.
Van der Voort, P.J. & Mol, H. (1989). *Basisdidactiek voor het onderwijs in de moderne vreemde talen.* Groningen: Wolters-Noordhoff.
Wenden, A. & Rubin, J. (1987). *Learning strategies in language learning.* Englewood Cliffs, N.J.: Prentice-Hall.
Westhoff, G.J. (1981). *Voorspellend lezen. Een didaktische benadering van de leesvaardigheidstraining in het moderne vreemde-talenonderwijs.* Groningen: Wolters-Noordhoff.
Willems, G. (1986). Communicatiestrategieën en hun rol in het vreemde-talenonderwijs. *Levende Talen 409*, 170-174.

Manuscript ontvangen 16-5-1994
Definitieve versie ontvangen 22-11-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 3, pp. 229-244

De rol van strategieën en metacognitie bij het leren van woorden in het moedertaal- en vreemde-talenonderwijs

Carolien Schouten-van Parreren[1], Kees de Glopper[2] en Maartje van Daalen-Kapteijns[2]
[1]*Vakgroep Pedagogiek, Vrije Universiteit Amsterdam**
[2]*SCO/Kohnstamm Instituut, Universiteit van Amsterdam*

ABSTRACT

Contextual inference, morphological analysis and estimation of word knowledge are three language learning strategies that can help enlarge students' vocabulary in their mother tongue as well as in foreign languages. The results of this study indicate that skill in the use of these strategies is substantially correlated with vocabulary knowledge, both within the mother tongue and within foreign languages. Also, between languages substantial correlations have been found between vocabulary knowledge and skill in word learning strategies. These outcomes support the assumption that training in the mother tongue may have beneficial effects for both mother tongue and foreign language vocabulary learning.

1. INLEIDING

Er zijn diverse strategieën om nieuwe woorden te leren. In dit artikel zal een onderzoek met betrekking tot drie van deze strategieën besproken worden. De bekendste van deze drie is het afleiden van de betekenis van een woord uit de context. Iets minder aandacht is er tot nu toe besteed aan het analyseren van woordvormen en nog vrijwel onbekend is de derde strategie: het schatten van de mate waarin de betekenis van een woord bekend is. Op deze laatste strategie zal in dit artikel het meest uitgebreid worden ingegaan. Overigens is deze strategie wel van een andere orde. Schatten van de eigen woordkennis kan namelijk nooit direct tot woordkennis leiden, maar alleen via het inzetten van andere strategieën, zoals afleiden en analyseren.

In deze paragraaf zullen we eerst aangeven op welke wijze het begrip strategie hier van toepassing is. Daarna gaan we in op de probleemstelling van het onderzoek.

De aandacht voor strategieën in het algemeen neemt de laatste tijd sterk toe, zoals Rijlaarsdam, Oostdam & Bimmel (elders in dit nummer) laten zien. Niet alleen prestaties, maar ook het 'leren leren' gaan steeds meer tellen in het talenonderwijs (Commissie Herziening Eindtermen 1990). Leerlingen moeten leren om strategieën bewust te hanteren, weten wanneer deze strategieën bruikbaar zijn en ze kunnen toepassen in situaties die enigermate verschillen van die waarin de strategieën geleerd zijn. Oostdam & Rijlaardam (1993, 1994) onderscheiden twee hoofdklassen talenstrategieën: taalgebruiksstrategieën met als hoofddoel het effectief gebruiken van taal voor communicatie in een functionele context en taalleerstrategieën met als hoofddoel het effectief leren van (aspecten van) een taal.

Als we dit onderscheid toepassen op de drie hierboven genoemde strategieën – afleiden, analyseren en schatten van de eigen woordkennis – dan gaat het in dit artikel vooral om taalleerstrategieën met het doel de eigen woordenschat uit te breiden. De drie strategieën kunnen overigens ook als taalgebruiksstrategieën ingezet worden ten aanzien van het taaldoel 'begrijpend lezen'. Voordat leerlingen in staat zijn om afleiden, analyseren en schatten als taalleer- of taalgebruiksstrategieën in te zetten, kan het bovendien nodig zijn ze deze strategieën expliciet te onderwijzen.

*Adres: Van der Boechorststraat 1, 1081 BT Amsterdam.

In dat laatste geval treedt de vaardigheid in het afleiden, analyseren en schatten op als leerdoel. Er kunnen dan weer andere leerstrategieën worden gehanteerd om deze leerdoelen te bereiken.

Ons onderzoek heeft vooral betrekking op afleiden, analyseren en schatten als mogelijke leerstrategieën ten aanzien van het leerdoel woorden leren. Hiermee sluit het aan bij een van de drie hoofddoelen die Graves (1987) voor het woordenschatonderwijs in de moedertaal onderscheidt: 'woorden leren', 'leren woorden leren' en 'leren over woorden'. Elk van deze doelen kent weer een aantal subdoelen. Aan 'woorden leren' wordt in het onderwijs de meeste aandacht besteed. Graves bespreekt in dit verband de volgende zes taken: het leren lezen van al bekende woorden (bij het technisch leren lezen), het leren van nieuwe betekenissen voor bekende woorden ('produkt' is óók: 'uitkomst van een vermenigvuldiging'), het leren van nieuwe woorden voor al bekende begrippen ('exploiteren' betekent 'gebruiken om er voordeel uit te halen'), het leren van nieuwe woorden voor nieuwe begrippen ('fascisme'), het verhelderen en verfijnen van de betekenis van al bekende woorden en het productief gebruiken van receptief gekende woorden. Gesteld al dat men dat zou willen, dan zijn er eenvoudigweg teveel woorden om ze allemaal gericht te onderwijzen.

Het tweede hoofddoel dat Graves voor het woordenschatonderwijs noemt, richt zich dan ook op het onderwijzen van strategieën die gericht zijn op het vergroten van de autonomie van de leerling bij het woorden leren, dus op het 'leren woorden leren'. Graves bespreekt er vijf: het gebruikmaken van de context, van de woordvorm en van woordenboeken, het ontwikkelen van een aanpak voor het omgaan met onbekende woorden in een tekst en het ontwikkelen van een persoonlijke aanpak voor het leren van woorden. Ons onderzoek heeft vooral betrekking op de eerste twee en ten dele op de vierde van deze strategieën.

Aan het derde hoofddoel dat Graves voor het woordenschatonderwijs van belang acht, het 'leren over woorden', wordt tot nog toe noch in het onderwijs noch in het onderzoek veel aandacht besteed. Graves denkt hierbij aan onderwerpen als 'leren wat het betekent om een woord te kennen', 'leren dat en hoe woordbetekenissen kunnen veranderen, 'leren om figuurlijk taalgebruik te herkennen en te gebruiken' e.a. Het eerste van deze onderwerpen, 'leren wat het betekent om een woord te kennen', heeft te maken met het schatten van de mate waarin de betekenis van een woord bekend is. Het gaat in ons onderzoek echter primair om een strategie en dat ligt dus meer op het terrein van Graves' tweede hoofddoel, het 'leren woorden leren'. Graves duidt deze strategie daar inderdaad (terloops) aan en merkt in dit verband al op dat 'recognizing that an unknown word has occurred – can be a big step for some students' (p. 177).

Het uiteindelijke doel van dit onderzoek is na te gaan in hoeverre het mogelijk is om ten aanzien van het woordenschatonderwijs transfer van het onderwijs in de moedertaal naar het onderwijs in vreemde talen te realiseren[1]. Hiertoe hebben we eerst onderzocht in welke mate woordkennis in de moedertaal samenhangt met woordkennis in vreemde talen en in welke mate woordkennis in de verschillende talen samenhangt met woordleerstrategieën en andere factoren. Over dit deel wordt hier verslag gedaan.

In de literatuur zijn diverse aanwijzingen te vinden voor de samenhang tussen woordkennis en woordleerstrategieën van verschillende aard. Beck & McKeown (1991) bespreken in een overzichtsartikel vier richtingen in het onderzoek over de wijze waarop de woordenschatverwerving in het moedertaalonderwijs het beste gestimuleerd zou kunnen worden. Het gaat hierbij om accentverschillen tussen de onderzoekers; de vier richtingen sluiten elkaar dus niet uit. De eerste richting gaat er vanuit dat leerlingen woorden vooral uit de context leren, bij het lezen. Maatregelen om te bevorderen dat leerlingen veel lezen zullen dus 'vanzelf' tot woordverwerving leiden (Nagy, Herman & Anderson 1985). De tweede richting ondersteunt het belang van veel lezen, maar wil leerlingen ook leren de betekenis van woorden uit de context af te leiden (Carnine, Kameenui & Coyle 1984). De derde richting zet daar vraagtekens bij en bepleit meer aandacht voor het gebruiken van een woordenboek (Schatz & Baldwin 1986). De vierde richting stelt dat directe instructie een belangrijke aanvulling is op het incidenteel leren van woorden, dat in de eerste drie richtingen wordt bepleit (Stahl & Fairbanks 1986).

Beck & McKeown (1991) zien in elk van de vier richtingen positieve aspecten, mits aan bepaalde voorwaarden is voldaan.

Het is de vraag of veel lezen voor het uitbreiden van de woordenschat voor *alle* leerlingen (dus niet alleen voor de goede) genoeg zoden aan de dijk zet. Training in het afleiden uit de context heeft tot bescheiden successen geleid (Jenkins, Matlock & Slocum 1989), maar er is meer onderzoek nodig om tot effectiever onderwijs te komen.

Woordenboekgebruik kan de woordenschatverwerving bevorderen, als je leerlingen zover kunt krijgen dat ze hun leesproces onderbreken om een woord op te zoeken en als de woordenboeken van dien aard zijn dat ze de betekenis van dat woord zonder moeite kunnen vinden. Reinking & Rickman (1990) vonden in dit verband dat leerlingen die een tekst lezen (op papier) met een computer onder handbereik om woorden op te zoeken, meer woorden opzoeken dan leerlingen die alleen een woordenlijst of woordenboek ter beschikking hebben. McKeown (1993) doet verslag van een onderzoek naar de effectiviteit van meer 'kindvriendelijke' woordenboekdefinities.

Tegenstanders van de directe instructie uit de vierde richting stellen dat het onmogelijk is om alle woorden direct te onderwijzen èn dat woordenschatonderwijs onnodig is: de leerlingen leren de woorden wel vanzelf. Het laatste argument gaat maar gedeeltelijk op. Er zijn in ieder geval zeer grote verschillen tussen leerlingen geconstateerd (Shefelbine 1990). Het eerste argument is waar, maar impliceert alleen dat woorden die in aanmerking komen voor gericht onderwijs, zorgvuldig geselecteerd moeten worden. McKeown & Beck (1988) gaan nader in op criteria voor een dergelijke selectie.

In het vreemde-talenonderwijs speelt directe instructie een vanzelfsprekende en grote rol, maar ook de andere drie stromingen zijn terug te vinden zijn in de literatuur over het vreemdetalenonderwijs. Krashen (1986) pleit voor veel lezen. Clarke & Nation (1981) en Van Parreren & Schouten-van Parreren (1981) pleiten voor een aanvullende training van het afleiden uit de context. Onderzoek naar de beperkingen hiervan is uitgevoerd door Mondria & Wit-de Boer (1991) en Hulstijn (1994).

In onderzoek ten aanzien van zowel het leren van woorden in de moedertaal als in vreemde talen speelt het afleiden van de betekenis van een woord dus een belangrijke, zij het niet onomstreden rol. Voor het afleiden zijn behalve de talige context ook de achtergrondkennis van de leerling en de woordvorm zelf van belang. Het analyseren van de woordvorm is impliciet aanwezig in het meeste onderzoek naar het afleiden (behalve als er niet met bestaande woorden, maar met pseudowoorden of met de cloze procedure – open plekken – wordt gewerkt). Er is empirische evidentie dat het analyseren van de woordvorm ('morphological generalization') ook een zelfstandige bijdrage levert aan de woordenschatuitbreiding (White, Power & White 1989). Deze bijdrage kan door gericht onderwijs nog vergroot worden (White, Sowell & Yanagihara 1989).

Het schatten van de eigen woordkennis wordt slechts incidenteel genoemd (Graves 1987), maar vormt wel een *voorwaarde* voor het succesvol in kunnen zetten van de strategieën 'afleiden' en 'analyseren'. Bij woorden die niet als onbekend worden gepercipieerd, zal immers ook geen poging worden ondernomen om de betekenis te achterhalen.

Er zijn dus wel aanwijzingen voor de samenhang tussen woordkennis en afleiden (Jenkins, Matlock & Slocum 1989), resp. tussen woordkennis en analyseren (White, Power & White 1989), maar er is nog nauwelijks systematisch onderzoek uitgevoerd naar de samenhang tussen de woordenschat in de moedertaal en in vreemde talen (Skehan 1989). Een sterke samenhang tussen woordkennis en woordleerstrategieën biedt geen uitsluitsel over de richting van het verband, maar geeft wel aanleiding om tot trainingsonderzoek over te gaan. Ook Palincsar & Brown (1984) volgen een dergelijke algemene onderzoeksstrategie. Bij hun onderzoek op het gebied van het begrijpend lezen hebben zij 1) op basis van de literatuur een theorie geformuleerd over de onderliggende cognitieve processen die leesvaardigheid kunnen beïnvloeden, 2) correlatieonderzoek uitgevoerd naar de samenhang tussen deze processen en leesvaardigheid en 3) een trainingsonderzoek uitgevoerd. Betere leesprestaties na een (adequate) training kunnen worden geïnterpreteerd als een ondersteuning van de theorie. Palincsar & Brown spreken in dit verband van een 'three-pronged approach to providing convergent evidence regarding the importance of any process as an underlying cognitive mechanism' (1984, p.170). Zoals al werd

opgemerkt, komt in dit artikel alleen het correlatieonderzoek aan de orde. Het trainingsonderzoek wordt op dit moment uitgevoerd.

Hieronder zullen eerst de vraagstelling, de opzet en de resultaten van het onderzoek beknopt aan de orde komen. Daarna zullen de onderzoeksvragen en resultaten ten aanzien van de nog betrekkelijk onbekende strategie van het schatten van de eigen woordkennis uitgebreid besproken worden. Om een zo breed mogelijk beeld te krijgen zijn in het onderzoek drie leeftijdsgroepen en drie talen opgenomen.

2. VRAAGSTELLING EN OPZET VAN HET ONDERZOEK

Het onderzoek richtte zich op de volgende vragen:

1. Hoe sterk is de samenhang tussen receptieve woordkennis in het Nederlands (als moedertaal), het Engels en het Frans van leerlingen van verschillende leeftijden?
2. Hoe sterk is de samenhang tussen relevante woordleerstrategieën en de receptieve woordkennis in het Nederlands, het Engels en het Frans van leerlingen van verschillende leeftijden?

Het onderzoek is uitgevoerd in groep 8 van het basisonderwijs (bo) en in leerjaar 2 en 4 van het voortgezet onderwijs (vo). Aan het onderzoek is deelgenomen door twintig basisscholen en door drie breed-heterogene scholengemeenschappen voor bavo/vbo-mavo-havo-vwo uit een middelgrote gemeente in het midden van het land. De samenstelling van de schoolbevolking vormt voorzover wij konden nagaan een redelijke afspiegeling van de totale Nederlandse leerlingenpopulatie, met als aantekening dat de proportie leerlingen met een vwo-advies of -opleiding enigszins onder het landelijk gemiddelde ligt. In elk van de drie leeftijdsgroepen hebben tussen de 350 en de 500 leerlingen aan het onderzoek deelgenomen[2].

De leerlingen hebben diverse toetsen gemaakt en een vragenlijst ingevuld. De leerlingen uit het bo hebben alleen toetsen voor het Nederlands gemaakt. De leerlingen uit het vo hebben ook toetsen voor het Engels en het Frans gemaakt. Het ging hierbij in de eerste plaats om toetsen voor woordkennis met daaraan gekoppeld een toets voor het schatten van de eigen woordkennis. Omdat het bij het schatten van de eigen kennis gaat om kennis over kennis kan hier van metacognitie gesproken worden. Om niet steeds de omslachtige term 'toets voor het schatten van eigen kennis' te hoeven gebruiken, zullen we in het vervolg van 'metacognitietoets' spreken. Daarnaast werden toetsen gemaakt voor het afleiden van de betekenis van een woord uit de context en morfologietoetsen voor het analyseren van de woordvorm (zie Bijlage 1 voor voorbeelditems). Toetsen van hetzelfde type waren zoveel mogelijk identiek voor de verschillende leeftijdsgroepen. Indien dat tot excessief lange toetsen zou leiden in verband met de grote range, bevatten de toetsen zoveel mogelijke gelijke of vergelijkbare items. Tenslotte zijn er door middel van een vragenlijst ook gegevens verzameld over andere factoren die blijkens de literatuur (Spolsky 1989) van invloed kunnen zijn, zoals het taalaanbod, de attitude en de motivatie van de leerlingen en diverse andere achtergrondkenmerken van de leerlingen. Deze komen in dit artikel niet uitvoerig aan de orde.

Woordkennis kan op verschillende manieren getoetst worden. Gezien de heterogeniteit van de leerlingen bij wie het onderzoek is uitgevoerd, kwam het er bijzonder op aan een toetsvorm te kiezen die zwakke leerlingen niet benadeelt. De toetsvorm waarbij de te toetsen woorden (onderstreept) worden aangeboden in een neutraal zinnetje, bleek hiervoor het meest geschikt te zijn (Van Wendel de Joode 1994). De korte zinnen werden gebruikt om de toets natuurlijker en aantrekkelijker te maken. De zinnen waren neutraal ten aanzien van de betekenis van het onderstreepte woord, omdat afleiden uit de context in deze toets geen rol mag spelen. Voor Engels en Frans werd ook de vertaling van de zinnen opgenomen. De leerlingen moesten de betekenis van het onderstreepte woord invullen op een open plaats in de vertaalde zin. Voor Nederlands is gebruik gemaakt van woordentoetsen met een vierkeuze vorm om te voorkomen dat verschillen tussen leerlingen in produktieve vaardigheid een te grote rol zouden spelen (bijv. bij het geven van de betekenis van woorden als 'inspireren' of 'ijdel'). Voor Engels en Frans was dat niet

nodig, omdat de leerlingen daar een vertaling in één woord kunnen geven en dus geen omschrijvingen hoeven te produceren.

Direct voorafgaand aan de corresponderende woordkennistoetsen zijn de metacognitietoetsen afgenomen. In deze toetsen is aan de leerlingen gevraagd op een vierpuntsschaal aan te geven in hoeverre ze de betekenis van een onderstreept woord in een kort zinnetje 'absoluut niet', 'waarschijnlijk niet', 'waarschijnlijk wel', 'zeker wel' zouden kunnen noemen. Het betrof hier steeds woorden die vervolgens ook in de woordkennistoets aan bod kwamen. Met de metacognitietoetsen wilden we meer inzicht verkrijgen in de relatie tussen het kunnen schatten van de eigen woordkennis en de werkelijk aanwezige woordkennis.

De afleidtoetsen waren opgebouwd uit een aantal korte tekstfragmenten met daarin één onbekend woord. De fragmenten waren voldoende pregnant om de betekenis van dit woord correct af te leiden. De fragmenten waren zo redundant dat het afleiden van het onderstreepte woord niet afhankelijk was van het niet kennen van de betekenis van een van de andere woorden in het fragment. De fragmenten zijn hierop gescreend met leerlingen van dezelfde leeftijd als de leerlingen voor wie de toetsen bedoeld waren. Als er maar enige twijfel bestond, werden woorden in de Engelse en Franse toetsen vervangen of indien dat niet mogelijk was, van een vertaling voorzien. Om de validiteit van de toetsen zo groot mogelijk te maken lieten we de leerlingen direct voor de afleidtoets een voortoets maken om na te gaan of ze de af te leiden woorden echt niet kenden. Als een leerling ook maar enig idee bleek te hebben over de betekenis van het af te leiden woord, bijv. 'iets met misdaad' voor 'cambrioler'(= inbreken), dan werd zo'n woord voor de betreffende leerling buiten de analyse gehouden.

Voor de Franse afleidtoets voor het tweede leerjaar ontwikkelden we een toets met geïllustreerde tekstfragmenten, omdat de woordkennis van leerlingen na één jaar onderwijs nog ontoereikend bleek te zijn voor het gebruiken van zuiver verbale fragmenten. Uit de analyse van hardopdenkprotocollen (Luttikhuis 1992) bleek dat leerlingen bij tekstfragmenten gecombineerd met plaatjes op eenzelfde wijze te werk gaan als bij tekstfragmenten zonder plaatjes.

Het analyseren van een woordvorm stelt de leerling voor een analyse- en een syntheseprobleem. De leerling moet een bekend woord (of een bekende stam) binnen een onbekend woord kunnen identificeren en de leerling moet op basis van de betekenis van dat woord (c.q. stam) en van de voor- en achtervoegsels de betekenis van het onbekende woord in zijn geheel kunnen bepalen (Wysocki & Jenkins 1987). In de morfologietoetsen voor de drie verschillende talen zijn voor beide aspecten toetsonderdelen opgenomen. Het eerste onderdeel bevat afgeleide woorden ('diskwalificeren') en niet-afgeleide woorden ('adresseren'). De leerling moet aangeven wat het grondwoord is (dus resp. 'kwalificeren' en 'adresseren'). Het tweede onderdeel bevat items die uit twee zinnetjes bestaan. In het eerste zinnetje wordt de betekenis van een grondwoord uitgelegd. In het tweede zinnetje moeten de leerlingen de betekenis invullen van een hiervan afgeleid woord. In dit artikel zijn alleen de resultaten op het tweede onderdeel verwerkt. De gegevens ten aanzien van het eerste toetsonderdeel worden nog nader geanalyseerd met behulp van hardop-denkprotocollen.

In tabel 1 zijn gegevens over de betrouwbaarheid van de verschillende toetsen samengevat. Uit de tabel blijkt dat de betrouwbaarheid van de toetsen voor woordkennis, afleiden en analyseren overwegend goed te noemen valt. Uitzonderingen op dit beeld zijn te vinden bij het vierde leerjaar Frans. Dit heeft wellicht te maken met het feit dat Frans in het vierde leerjaar van de onderzochte scholen een bijzondere positie inneemt. Het vak wordt niet meer door alle leerlingen gevolgd. De leerlingen die Frans hebben gekozen, vormen een heterogene groep: er zijn zowel positieve kiezers bij als negatieve kiezers, dat wil zeggen leerlingen die Frans kiezen, omdat ze andere vakken nog moeilijker of minder aantrekkelijk vinden. Uit de tabel valt verder op te maken dat de betrouwbaarheid van de metacognitietoets te wensen overlaat: de overschattingsscores zijn overwegend matig betrouwbaar.

Tabel 1. Betrouwbaarheid (Cronbach alpha) van de toetsen voor woordkennis, afleiden en analyseren, aantal items en aantal waarnemingen waarop betrouwbaarheidsschatting gebaseerd is (k/N).

	Nederlands			Engels		Frans	
	gr 8	lj 2	lj 4	lj 2	lj 4	lj 2	lj 4
Woordkennis	.90 [64/329]	.89 [60/402]	.90 [64/256]	.94 [60/304]	.96 [59/282]	.93 [50/324]	.90 [66/94]
Overschatten	.59 [30/340]	.51 [30/240]	.47 [30/303]	.69 [30/320]	.66 [30/272]	.76 [30/296]	.37 [30/89]
Afleiden	.78 [36/298]	.78 [36/224]	.82 [36/151]	.93 [37/145]	.94 [43/.94]	.85 [36/50]	.72 [34/56]
Analyseren	.85 [30/424]	.84 [30/306]	.86 [30/297]	.83 [28/289]	.86 [28/278]	.73 [33/300]	.67 [27/88]

3. RESULTATEN

Alvorens de resultaten voor de samenhang tussen het schatten van de eigen woordkennis en de daadwerkelijke woordkennis te bespreken, beantwoorden we eerst de twee algemene onderzoeksvragen.

In tabel 2 zijn de samenhangen tussen woordkennis in Nederlands, Engels en Frans voor het tweede en het vierde leerjaar weergegeven. Hiermee kan de eerste onderzoeksvraag beantwoord worden. De samenhangen tussen woordkennis in het Nederlands en het Engels blijken het sterkst te zijn, gevolgd door die tussen het Nederlands en het Frans. De samenhangen tussen het Engels en het Frans zijn matig tot zwak (correlaties tussen .00 en .20 worden zwak genoemd, correlaties tussen de .20 en .40 matig en correlaties groter dan .40 sterk, Cohen 1988). De lage correlaties voor Frans in het vierde leerjaar kunnen (mede) te wijten zijn aan de al eerder vermelde bijzondere positie van dit vak in dit leerjaar. Het is niet waarschijnlijk dat de gevonden correlaties vooral toegeschreven moeten worden aan het feit dat woordkennis in de ene taal woordkennis in de andere taal direct beïnvloedt, al zal dit wel een zekere rol spelen. In kwalitatief onderzoek bleek bijvoorbeeld dat zwakke leerlingen in tegenstelling tot sterke leerlingen (zwak-sterk wat betreft scores op woordenschattoetsen en het oordeel van de vakdocent) bepaalde Nederlandse woorden die steun kunnen bieden bij het leren van de betekenissen van vreemdtalige woorden, niet kennen. Deze leerlingen hadden bijvoorbeeld nog nooit van een woord als 'arriveren' gehoord, waardoor de betekenis 'aankomen' van 'to arrive' of 'arriver' voor hen geheel nieuw en moeilijk te leren is, terwijl het voor sterke leerlingen juist om verwan-

Tabel 2. Produkt-moment correlaties tussen woordkennisscores voor Nederlands, Engels en Frans (tussen haken 95%-betrouwbaarheidsinterval).

	Engels		Frans	
	lj 2	lj 4	lj 2	lj 4
Nederlands	.47 [.38-.55]	.58 [.49-.65]	.42 [33-.51]	.18 [-.03-.37]
Engels	*	*	.33 [.23-.42]	.13 [-.09-.34]

Tabel 3. Produkt-moment correlaties tussen scores voor afleidvaardigheid en analysevaardigheid voor Nederlands, Engels en Frans (tussen haken 95%-betrouwbaarheidsinterval).

	Afleiden			
	Engels		Frans	
	lj 2	lj 4	lj 2	lj 4
Nederlands	.56 [.48-.63]	.63 [.55-.70]	.57 [.49-.65]	.49 [.31-.64]
Engels	*	*	.58 [.49-.65]	.41 [.21-.59]
	Analyseren			
	Engels		Frans	
	lj 2	lj 4	lj 2	lj 4
Nederlands	.59 [.51-.67]	.66 [.59-.72]	.39 [.28-.49]	.38 [.18-.73]
Engels	*	*	.44 [.34-.53]	.44 [.24-.60]

te, dus gemakkelijk te leren woorden gaat (Schouten-van Parreren 1990). De gevonden correlaties zullen dus waarschijnlijk grotendeels toegeschreven moeten worden aan andere factoren dan de directe beïnvloeding van woordkennis in de ene taal óp die in de andere taal. Deze factoren zullen nu aan de orde komen, bij de beantwoording van de tweede onderzoeksvraag.

Zoals uit tabel 3 is af te lezen, is het aannemelijk dat tot deze factoren de woordleerstrategieën behoren. Zowel het afleiden als het analyseren vertonen sterke samenhangen tussen talen. Het feit dat leerlingen die in het Nederlands goed kunnen afleiden en analyseren dat overwegend ook in het Engels en Frans goed doen, zou dus voor een deel een verklaring kunnen vormen voor de tamelijk sterke samenhangen tussen woordkennis in de verschillende talen. De resultaten ten aanzien van het schatten van de eigen woordkennis komen in paragraaf 5 aan de orde.

In tabel 4 is te zien dat de correlaties tussen afleiden en analyseren enerzijds en woordkennis anderzijds voor alle taal-leeftijdsgroepen sterk zijn. Dit ondersteunt het belang van deze woordleerstrategieën voor het uitbreiden van de woordenschat. De samenhangen tussen woordkennis

Tabel 4. Produkt-moment correlaties tussen woordkennisscores en scores voor afleidvaardigheid en analysevaardigheid voor Nederlands, Engels en Frans (tussen haken 95%-betrouwbaarheidsinterval).

	Woordkennis						
	Nederlands			Engels		Frans	
	gr 8	lj 2	lj 4	lj 2	lj 4	lj 2	lj 4
Afleiden (resp. NEF)	.69 [.63-.74]	.62 [.55-.69]	.64 [.57-.70]	.64 [.57-.70]	.51 [.41-.59]	.58 [.49-.65]	.65 [.51-.77]
Analyseren (resp. NEF)	.71 [.66-.76]	.64 [.57-.70]	.67 [.60-.73]	.45 [.35-.54]	.56 [.47-.65]	.40 [.29-.49]	.50 [.32-.65]

en woordleerstrategieën zijn veel sterker dan die tussen woordkennis en de overige door middel van de vragenlijst gemeten variabelen (zoals aanbod, motivatie, attitude). Hierop zal in een ander artikel nader worden ingegaan. Aangezien de samenhangen tussen de talen sterk zijn, lijkt transfer naar de vreemde talen bij training van woordleervaardigheden in de moedertaal kansrijk.

We zullen nu ingaan op de onderzoeksvragen en resultaten ten aanzien van het schatten van de eigen woordkennis.

4. SCHATTEN VAN DE EIGEN WOORDKENNIS – EEN KWESTIE VAN METACOGNITIE

Als leerlingen stelselmatig denken woorden wel te kennen, terwijl dat niet het geval is, zullen ze er bij het leren geen aandacht meer aan besteden. Dit staat een uitbreiding van hun woordenschat in de weg. Zo luidt het probleem waar we hier nader op ingaan. Hoewel er steeds meer onderzoek wordt uitgevoerd naar de rol van metacognitie bij het leren, is er nog maar sporadisch onderzoek beschikbaar dat relevant is voor (bepaalde aspecten van) de woordenschatverwerving (Anderson & Freebody 1983, Nelson & Dunlosky 1991, Zechmeister et al. 1993). Om meer inzicht te krijgen in de relatie tussen het kunnen schatten van de eigen woordkennis en het daadwerkelijk kennen van woorden hebben we de vraag naar de mate van samenhang tussen deze beide factoren nader uitgewerkt in de volgende vier deelvragen:

1. In hoeverre zijn leerlingen van verschillende leeftijden in staat om hun woordkennis goed in te schatten in de moedertaal en in vreemde talen?
2. Zijn zwakke leerlingen meer geneigd om hun woordkennis te overschatten dan sterke leerlingen?
3. Is overschatten taalonafhankelijk (dat wil zeggen overschatten dezelfde leerlingen hun woordkennis in verschillende talen)?
4. Zijn er specifieke woorden die vaker overschat worden dan andere?

Deze vragen zullen in de volgende paragraaf beantwoord worden.

5. RESULTATEN VOOR METACOGNITIE

5.1 Leeftijdsverschillen?

Om na te gaan in hoeverre leerlingen van verschillende leeftijden in staat zijn om hun woordkennis in te schatten zijn per taal en per leeftijdsgroep produkt-moment correlaties berekend tussen de p-waarden van de items uit de woordkennistoetsen en de gemiddelde scores op de vierpuntsschaal voor de items uit de metacognitietoets. In tabel 5 zijn deze correlaties tussen de gemiddelde geschatte en daadwerkelijke woordkennis weergegeven. De correlaties stijgen naarmate de leerlingen ouder worden. Dit geldt zowel voor het Nederlands als voor het Engels en het Frans. Waardoor stijgen deze correlaties? De meest voor de hand liggende verklaring is dat leerlingen naarmate zij ouder worden, steeds beter in staat zijn om hun kennis goed in te schatten. Om steun voor deze veronderstelling te vinden hebben we gekeken naar het percentage overschatte woorden voor de verschillende leeftijden en talen. Als overschat hebben we die woorden beschouwd die op de metacognitietoets een '4' kregen ('ik kan de betekenis zeker wel noemen'), terwijl het corresponderende antwoord op de woordkennistoets fout was. Het percentage overschatte woorden is bepaald als het aantal overschatte woorden op het aantal niet gekende woorden. In tabel 6 zien we dat een afname van het percentage overschattingen inderdaad een oorzaak kan zijn voor de toenemende samenhang tussen ingeschatte en daadwerkelijke woordkennis. De afname in het percentage overschatte woorden geldt voor alle drie talen.

Het is dus waarschijnlijk dat oudere leerlingen beter in staat zijn om de eigen woordkennis goed in te schatten, maar hoe komt dat? Er zijn verschillende verklaringen denkbaar:

Tabel 5. Produkt-moment correlaties tussen p-waarden van de items uit de woordkennistoetsen (feitelijke woordkennis) en de gemiddelde scores op de de items uit de metacognitietoets (geschatte woordkennis) voor Nederlands, Engels en Frans (N is steeds 30 woorden).

	gr 8	lj 2	lj 4
Nederlands	.50	.66	.77
Engels	*	.67	.81
Frans	*	.77	.80

Tabel 6. Percentage overschatte woorden van het totaal aantal niet gekende woorden.

	Nederlands			Engels		Frans	
	gr 8	lj 2	lj 4	lj 2	lj 4	lj 2	lj 4
Gemiddelde	28.6	24.6	14.7	42.6	34.7	38.7	23.8
SD	18.7	18.4	15.8	28.3	25.9	27.3	17.8
N	455	336	345	319	281	316	94

- De oudere leerlingen vormen een meer selecte groep. Dit geldt inderdaad voor Frans in het vierde leerjaar, maar nauwelijks voor de andere groepen, aangezien het onderzoek is uitgevoerd in breed-heterogene scholen. Deze verklaring is dus slechts van ondergeschikt belang.
- De leeftijd zelf speelt een rol. In de literatuur over metacognitie op andere gebieden blijkt dat de metacognitieve kennis en vaardigheden van kinderen zich ontwikkelen naarmate zij ouder worden (Garner 1988, Neuvel, Bossers, Schoonen & Hulstijn 1993). De leeftijdsfactor kan dus van belang zijn.
- In tabel 6 kunnen we zien, dat leeftijd niet de enige factor is die de stijgende correlatie tussen ingeschatte en daadwerkelijke woordkennis kan verklaren. Waarschijnlijk is ook de fase van het taalleerproces waarin de leerling verkeert, van belang. Leerlingen die een nieuwe taal beginnen te leren, worden bijvoorbeeld met andere categorieën van woorden geconfronteerd dan het geval is in een taal waarmee zij al meer vertrouwd zijn. We komen hierop nog terug in 5.4.

De algemene conclusie kan dus luiden dat oudere leerlingen beter in staat lijken tot het goed kunnen inschatten van hun woordkennis. Het juist kunnen inschatten van de eigen woordkennis lijkt ook af te hangen van de fase van het taalleerproces waarin leerlingen verkeren.

5.2 Niveauverschillen?

In tabel 6 zien we dat leerlingen in alle talen en leeftijdsgroepen hun woordkennis in percentueel opzicht geregeld tot zeer vaak overschatten. Bij Nederlands in leerjaar 4 wordt gemiddeld 14.7% van de niet gekende woorden overschat, bij Engels in het tweede leerjaar zelfs 42.6%. In absoluut opzicht valt het aantal keren dat leerlingen hun woordkennis overschatten overigens wel mee. Het gemiddeld aantal overschatte woorden varieert tussen 1.64 (Nederlands in leerjaar 4) en 4.33 (Frans in leerjaar 2).

In kleinschalig kwalitatief onderzoek met zwakke en sterke leerlingen was eerder naar voren gekomen dat een aantal van de zwakke leerlingen de eigen woordkennis in extreme mate overschat. In deze protocollen waren overigens geen aanwijzingen te vinden voor het voorkomen van onderschatting. Er waren nog twee andere redenen om onderschatten niet in dit onderzoek te betrekken. Ten eerste is de didactische relevantie van onderschatten geringer dan die van overschatten. Ten tweede zou het evalueren van onderschatting een aanpassing van de door ons

ontwikkelde toetsen en het afnemen van extra toetsen (bijv. voor faalangst) vereisen. Hier volgen enkele voorbeelden uit de protocollen van overschatte woorden[3] ('4' op de metacognitietoets/fout op de woordkennistoets):

quinze
pp: (schrijft 'twaalf' op) Toch? Nee, nee, nee .. die is het.
pl: Wat heb je er nu van gemaakt?
pp: Dertien (!). Ik moest even denken.

piscine
pp: Piscine! Die heb ik gehad.
pl: Mm mm.
pl: .. Ja, daar twijfel ik over.
pl: Ja, waar twijfel je over?
pp: Nou, een kip of vis (!).
pl: Ja. Maar je zet hem toch maar op 4? Want als je het niet precies weet ..
pp: ja, dat weet ik .. ja..
pl: dan zou je hem eigenlijk bij 3 zetten, weet je nog?
pp: Ja, dat weet ik.
pl: Maar je vindt hem toch eh..
pp: Ja. Ik weet hem wel.

afzondering
pl: En dan afzondering?
pp: Afzondering, eh .., afzondering .. uitzondering weet ik wel.
pl: Ja, maar dat staat er niet.
pp: Nee, dat weet ik. Afzondering zal wel het tegengestelde van uitzondering zijn.
pp: Leg dan eens uit wat afzondering zou moeten betekenen?
pp: Afzondering is dat je erbij hoort.

In het onderhavige onderzoek zijn we nagegaan in hoeverre we een bevestiging konden vinden voor het bestaan van een groep zwakke leerlingen die de eigen woordkennis in hoge mate overschat. Leerlingen met beperkte woordkennis blijken in absoluut opzicht hun woordkennis inderdaad vaker te overschatten. Uit tabel 7 valt op te maken dat het steeds de zwakke leerlingen zijn, die vaker aangeven dat zij een woord kennen, terwijl zij op de toets het goede antwoord schuldig blijven. De enige uitzondering vormt weer Frans in het vierde leerjaar.

Dat zwakke leerlingen vaker overschatten is op zichzelf niet zo verbazend, omdat deze leerlingen in de toets meer onbekende woorden tegenkomen en dus meer gelegenheden tot het overschatten van hun kennis hebben. Verschillen in kansen op overschatten doen niet af aan de betekenis van het resultaat dat zwakke leerlingen in absolute zin vaker overschatten. Niet alleen in de onderhavige toetssituatie, maar ook in andere situaties (m.n. in het talenonderwijs) zullen zwakke leerlingen vaker in aanraking komen met onbekende woorden.

Met verschillen in kansen op overschatten valt overigens wel rekening te houden door niet het absolute aantal maar het percentage overschatte woorden in ogenschouw te nemen. Zijn zwakke leerlingen ook percentueel vaker geneigd een woord dat ze niet kennen te overschatten? Dat is niet het geval. Het beeld lijkt eerder de andere richting uit te wijzen: naarmate leerlingen meer woorden kennen, zijn zij meer geneigd zijn tot het overschatten van woorden die ze niet kennen. Nader onderzoek zou moeten uitwijzen hoe dit te verklaren is. In tabel 7 zijn de resultaten inzake het absolute en het percentuele overschatten samengevat.

Dat er een kleine groep van extreme overschatters is, leerlingen die bij veel woorden hun kennis overschatten, wordt geïllustreerd in figuur 1. In dit histogram voor de verdeling van het aantal overschatte woorden bij Frans in het tweede leerjaar valt een kleine 'staart' van leerlingen te ontwaren die tien of meer van de dertig aangeboden woorden overschat. De verdelingen voor de andere taal- en leeftijdsgroepen vertonen sterke gelijkenis met die in figuur 1.

In absoluut opzicht overschatten de meeste leerlingen de eigen woordkennis niet erg vaak,

Tabel 7. Gemiddelde waarden voor absoluut aantal overschattingen en voor percentage overschatte woorden voor tien woordkennis-decielgroepen (groep 1=minste woordkennis, groep 10=meeste woordkennis).

Taal/ Leeftijd	Decielgroep Woordkennis 1	2	3	4	5	6	7	8	9	10	df	F	p
Ned gr8: abs	5.5	5.0	4.8	4.1	3.8	4.2	3.1	2.7	1.8	1.0	9	19.1	.00
Ned gr8: %	26	26	29	25	27	31	27	31	31	32	9	1.0	.46
Ned lj2: abs	3.9	3.5	3.2	3.0	2.9	2.8	2.7	1.2	1.7	1.0	9	10.5	.00
Ned lj2: %	21	21	22	24	24	28	33	17	31	28	9	2.6	.00
Ned lj4: abs	2.1	1.8	2.2	2.1	1.7	1.5	1.6	1.4	0.9	1.0	9	2.4	.01
Ned lj4: %	10	10	14	15	13	14	16	16	15	22	9	1.7	.10
Eng lj2: abs	6.5	5.7	4.7	3.8	2.9	2.6	2.3	1.6	1.1	0.3	9	29.2	.00
Eng lj2: %	31	38	40	43	42	44	51	47	55	33	9	2.1	.03
Eng lj4: abs	4.6	6.1	4.9	5.2	4.6	3.4	3.2	2.4	1.8	1.1	9	10.5	.00
Eng lj4: %	18	28	26	34	37	33	35	36	40	61	9	5.9	.00
Fra lj2: abs	5.8	6.1	5.0	5.4	5.6	4.8	4.3	3.5	1.9	0.9	9	8.8	.00
Fra lj2: %	23	31	28	37	40	42	44	47	42	59	9	4.6	.00
Fra lj4: abs	1.6	2.4	2.6	2.2	2.8	2.8	3.3	2.2	1.4	1.3	9	1.3	.23
Fra lj4: %	8	14	18	21	27	28	32	30	27	31	9	2.1	.04

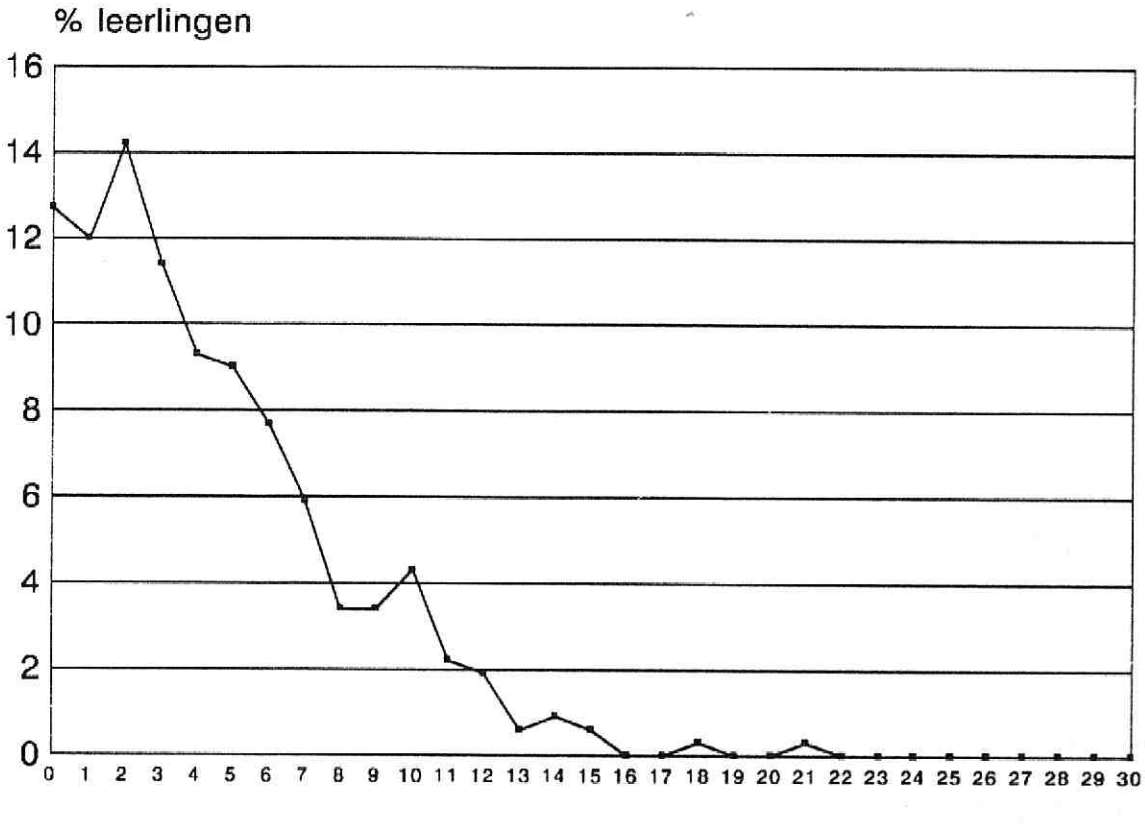


Fig. 1. Scoreverdeling overschatten voor Frans, leerjaar 2.

maar een kleine groep van overwegend zwakke leerlingen vertoont dat gedrag in zo'n extreme mate dat het zeker zinvol is hier aandacht aan te besteden.

5.3 Taalverschillen?

Om na te gaan of overschatten taalspecifiek of taalonafhankelijk is, hebben we de gegevens van de leerlingen van het tweede leerjaar vo geanalyseerd. We hebben ons beperkt tot het tweede leerjaar, omdat in de scholen die aan het onderzoek deelnamen, deze groep leerlingen in zijn geheel verplicht onderwijs volgt in alle drie de talen. Er is sprake van zeer scheve verdelingen: de meeste leerlingen overschatten af en toe, terwijl een kleine groep leerlingen extreem vaak overschat. Bij onze analyses van de wijze waarop leerlingen over verschillende schattingsgroepen verdeeld zijn, namen we als uitgangspunt dat niet meer dan 25% van de leerlingen per taalleeftijdsgroep tot de extreme overschatters behoorden (uit de histogrammen blijkt dat de groep extreme overschatters zeker niet groter is). De kruistabelanalyses in tabel 8 laten zien dat het extreem overschatten niet louter taalspecifiek is. Voor ieder van de drie combinaties (Nederlands-Engels, Nederlands-Frans en Engels-Frans) is de chi-kwadraat waarde significant op het .05 niveau. De geobserveerde verdeling wijkt steeds het sterkst af van de toevalsverdeling in de cellen rechtsonder. Leerlingen die in één taal sterk overschatten hebben een verhoogde kans dat ook in een andere taal te doen. Sterk is de samenhang tussen het overschatten in verschillende talen overigens niet. De waarden van coëfficiënt phi liggen liggen tussen .12 en .15.

Er kan de volgende conclusie getrokken worden. Het extreem overschatten van de eigen woordkennis is niet helemaal taalonafhankelijk. Leerlingen die in één taal extreem overschatten hebben een verhoogde kans dat ook in een andere taal te doen.

5.4 Woordverschillen?

Uit het kwalitatieve vooronderzoek waren aanwijzingen naar voren gekomen dat bepaalde categorieën van woorden eerder overschat worden dan andere. In dit onderzoek werden deze eerdere bevindingen grotendeels bevestigd en verder verfijnd. Hiertoe werden 'overschattingsranglijsten' voor de woorden gemaakt en werden de antwoorden van de extreme overschatters in kaart gebracht (voor het tweede leerjaar). De volgende drie categorieën van woorden blijken het meest aanleiding te geven tot overschatting:

1. Woorden uit een specifieke groep, zoals telwoorden, vraagwoorden, weekdagen, kledingstukken, kleuren. Deze categorie is vooral frequent bij het Frans waar veel woorden (nog) geen enkel houvast bieden bij het leren, terwijl ze in het onderwijs bovendien vaak als groep worden aangeboden.
2. Woorden die veel op andere woorden in de moedertaal (of in de vreemde taal) lijken, zoals 'gate-gat' en 'snakes-snacks'. Deze categorie komt vooral veel bij het Engels voor waar leerlingen al veel woorden denken te (her)kennen, voor een deel dus ten onrechte.
3. Woorden die vertrouwd klinken, terwijl hun vaak vrij gecompliceerde betekenis niet bekend is. Dit komt vooral voor in de moedertaal. Sommige leerlingen hebben woorden als 'tole-

Tabel 8. Overschatting in Nederlands, Engels en Frans: verdeling van leerlingen uit leerjaar 2 over cellen van tweewegs-kruistabellen (0=overschat niet sterk in N, E of F; 1=overschat sterk in N, E of F).

		Eng 0	Eng 1
Ned	0	199	54
	1	60	30
χ^2=4.53 df=1 p=.03 phi=.12			

		Fra 0	Fra 1
Ned	0	186	62
	1	54	36
χ^2=6.51 df=1 p=.01 phi=.15			

		Fra 0	Fra 1
Eng	0	185	63
	1	52	32
χ^2=4.35 df=1 p=.04 phi=.12			

rant', 'actueel' of 'economisch' al zo vaak gehoord dat ze denken ook de betekenis te kennen, terwijl dat geenszins het geval blijkt te zijn.

Hoewel alle leerlingen bij woorden van deze drie categorieën meer overschatten, lijken vooral zwakke leerlingen geen idee te hebben van het dreigende interferentiegevaar bij woorden die op elkaar lijken qua betekenis en/of qua klank (categorie 1 en 2). Ook lijken zwakke leerlingen zich onvoldoende bewust te zijn van het gegeven dat een woord wèl vertrouwd kan klinken zonder dat de betekenis bekend is. Dit zou samen kunnen hangen met het feit dat de kwaliteit van de woordbetekenissen waarover zwakke leerlingen beschikken in het algemeen relatief laag is (Elshout-Mohr & Van Daalen-Kapteijns 1987). Van belang voor het onderwijs is dat het mogelijk blijkt te zijn om te voorspellen welke woorden leerlingen zullen overschatten, zowel voor de moedertaal als voor vreemde talen. In Schouten-van Parreren, De Glopper & Van Daalen-Kapteijns (1993) zijn concrete suggesties gedaan voor didactische maatregelen die genomen kunnen worden om het overschatten van de drie categorieën woorden zoveel mogelijk tegen te gaan.

6. CONCLUSIES

Woordkennis in de moedertaal en in vreemde talen hangt samen. De samenhang kan voor een aanzienlijk deel toegeschreven worden aan de mate waarin de leerlingen over de woordleerstrategieën afleiden, analyseren en schatten van de eigen kennis beschikken. Deze strategieën hangen namelijk stuk voor stuk *binnen* de talen sterk samen met de woordkennis (tabel 4 en 5) en *tussen* de talen sterk met elkaar (tabel 3). Deze bevinding onderstreept het belang van training van de woordleerstrategieën in het onderwijs. Bij het afleiden (Van Daalen-Kapteijns, Schouten-van Parreren & De Glopper 1993) en bij het analyseren (Van Daalen-Kapteijns, De Glopper, Schouten-van Parreren & Baak 1993) is hierbij voor een grote groep leerlingen nog winst te boeken. Bij het schatten van de eigen woordkennis gaat het, zoals we in par. 5 zagen, om een kleinere groep leerlingen, die de eigen kennis echter wel in hoge mate overschat. Dit overschatten staat de woordenschatuitbreiding van deze leerlingen dusdanig in de weg dat ook hier gerichte maatregelen tot een aanzienlijke verbetering zouden kunnen leiden.

Gezien de sterke correlaties tussen de woordleerstrategieën in de moedertaal en de vreemde talen lijkt transfer tussen de talen kansrijk. Training in woordleerstrategieën in de moedertaal zou dan ook profijt af kunnen werpen voor de vreemde talen. Als leerlingen zich bijvoorbeeld meer bewust zouden worden van het gevaar om bepaalde woorden te overschatten in de moedertaal, zou dat inzicht ook een nuttige rol kunnen spelen bij het leren van woorden in de vreemde talen.

Onderzoek naar de effectiviteit van training in woordleerstrategieën en naar de voorwaarden waaronder transfer plaatsvindt, is onlangs van start gegaan. Bij dit trainingsonderzoek zal rekening gehouden worden met de door Rijlaarsdam, Oostdam & Bimmel (elders in dit nummer) beschreven componenten die strategietraining effectief en kansrijk voor transfer kunnen maken: leerlingen moeten gaan beschikken over strategieën die ze bewust kunnen hanteren en ze moeten weten wanneer ze deze wel en niet in kunnen zetten. Reflecteren op strategiegebruik kan leerlingen helpen het eigen gedrag steeds meer te reguleren.

NOTEN

1. Dit onderzoek maakt deel uit van een Programma voor Strategisch Onderzoek (PSO, gefinancierd door SVO) dat wordt uitgevoerd door het SCO-Kohnstamm Instituut in samenwerking met de universitaire lerarenopleidingen van de RUU (IVLOS), de UvA (ILO) en de VU (IDO) en met de vakgroep Taalkunde van de VU. Het onderzoek richt zich op leerganganalyse, leesvaardigheid en woordenschatverwerving. In dit artikel komt alleen dit laatste onderdeel aan de orde (zie voor een overzicht van het PSO in zijn geheel: Levende Talen, dec. 1993; De Glopper, Schouten-van Parreren & Van Daalen-Kapteijns 1993 geven een overzicht van het onderdeel 'woordenschatverwerving').

2. Hierbij danken wij de betrokken leerlingen, leraren en scholen hartelijk voor hun medewerking aan het onderzoek.
3. Het hardopdenk-onderzoek voor Nederlands is uitgevoerd door Ben van Wendel de Joode.

LITERATUUR

Anderson, R.C. & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Hutton (ed.). *Advances in reading/language research: A research annual* (p. 231-256). Greenwich, CT: JAI Press.

Beck, I. & McKeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (eds.). *Second handbook of reading research* (p. 789-814). New York: Longman.

Carnine, D., Kameenui, E.J., & Coyle, G. (1984). Utilization of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly, 19*, 188-204.

Clarke, D.F. & Nation, I.S.P. (1980). Guessing the meanings of words from context: Strategy and techniques. *System 8*, 211-220.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum.

Commissie Herziening Eindtermen (1990). *Kerndoelen voor de basisvorming in basisonderwijs en voortgezet onderwijs*. Zoetermeer: O&W/DOP.

Daalen-Kapteijns, M. van, Schouten-van Parreren, M.C., & Glopper, K. de (1993). Het afleiden van woordbetekenissen uit de context. *Levende Talen, 485*, 589-593.

Daalen-Kapteijns, M. van, Schouten-van Parreren, M.C., Glopper, K. de, & Baak, E. (1993). Het analyseren van woordvormen en kennis van voorvoegsels. *Levende Talen, 485*, 594-597.

Elshout-Mohr, M. & Daalen-Kapteijns, M. van (1987). Cognitive processes in learning word meanings. In M.G. McKeown & M.E. Curtis (eds.). *The nature of vocabulary acquisition* (p. 53-71). Hillsdale, N.J.: Lawrence Erlbaum.

Garner, R. (1988). *Metacognition and reading comprehension*. Norwood, N.J.: Ablex.

Glopper, K. de, Schouten-van Parreren, M.C., & Daalen-Kapteijns, M. van (1993). Woordenschat en woordleervaardigheden in moedertaal en vreemde talen. *Levende Talen, 485*, 584-588.

Graves, M.F. (1987). The roles of instruction in fostering vocabulary development. In: M.G. McKeown, & M.E. Curtis (eds.). *The nature of vocabulary acquisition* (p. 165-184). Hillsdale, N.J.: Lawrence Erlbaum.

Hulstijn, J.H. (1994). L'acquisition incidente du lexique en langue étrangère au cours de la lecture: ses avantages et ses limites. *Acquisition et Interaction en Langue Étrangère 3*, 77-96.

Jenkins, J.R., Matlock, B., & Slocum, T.A. (1989). Two approaches to vocabulary instruction: the teaching of individual word meanings and practice in deriving word meaning from context. *Reading Research Quarterly, 24*, 215-235.

Krashen, S.D. (1986). We acquire vocabulary by reading. In A. Papalia (ed.). *Teaching our students a second language in a proficiency-based classroom*. New York: New York State Association of Foreign Language Teachers.

Luttikhuis, Y. (1992). *L'image dans un test de déduction contextuelle*. Amsterdam: Vrije Universiteit (doctoraal scriptie).

McKeown, M.G. (1993). Creating effective definitions for young word learners. *Reading Research Quarterly, 28*, 16-31.

McKeown, M.G. & Beck, I.L. (1988). Learning vocabulary: different ways for different goals. *Remedial and Special Education, 9*, 42-52.

Mondria, J. & Wit-de Boer M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics, 12*, 249-267.

Nagy, W.E., Herman, P.A., & Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly, 20*, 233-53.

Nelson, T.O. & Dunlosky, J. (1991). When peoples judgements of learning (JOLs) are extremely accurate at predicting subsequent recall: The "Delayed-JOL Effect. *American Psychological Society, 2*, 267-270.

Neuvel, J., Bossers, B., Schoonen, R., & Hulstijn, J. (1993). *Levende Talen, 485*, 603-610.

Oostdam, R. (Ed.) (1993). Transfer in het talenonderwijs. *Levende Talen, 485*, (themanummer transfer).

Oostdam, R. & Rijlaardam, G. (1993). Strategisch georiënteerd moedertaalonderwijs. *Spiegel 11*, 9-30.

Oostdam, R. & Rijlaarsdam, G. (1994). *Towards Strategic Language Learning*. Amsterdam: SCO/ILO.
Palincsar, A.S. & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117-175.
Parreren, C.F. van & Schouten-van Parreren, M.C. (1981). Contextual guessing: a trainable reader strategy. *System 9*, 235-241.
Reinking, D. & Rickman, S.S. (1990). The effects of computer-mediated texts on the vocabulary learning and comprehension of intermediate-grade readers. *Journal of Reading Behavior, 22*, 395-410.
Schatz, E.K. & Baldwin, R.S. (1986). Context clues are unreliable predictors of word meanings. *Reading Research Quarterly, 21*, 439-453.
Schouten-van Parreren, M.C. (1992). Individual differences in vocabulary acquisition. In P.J.L. Arnaud & H. Béjoint (eds.). *Vocabulary and applied linguistics* (pp. 94-101). London: MacMillan.
Schouten-van Parreren, M.C., Glopper, K. de & Daalen-Kapteijns, M. van (1993). Het overschatten van de eigen woordkennis. *Levende Talen, 485*, 599-602.
Shefelbine, J.L. (1990). Students factors related to variability in learning word meanings from context. *Journal of Reading Behavior, 22*, 71-98.
Skehan, P. (1989). *Individual differences in second-language learning*. London: Edward Arnold.
Spolsky, B. (1989). *Conditions for Second Language Learning. Introduction to a general theory*. Oxford: Oxford University Press.
Stahl, S.A. & Fairbanks, M.M. (1986). The effects of vocabulary instruction: a model-based meta-analysis. *Review of Educational Research, 56*, 72-110.
Wendel de Joode, B. van (1994). Tests de vocabulaire néerlandais pour élèves forts et élèves faibles. In F.G. Brinkman, J.A. van der Schee & M.C. Schouten-van Parreren (eds.). *Curriculum research different disciplines and common goals. The development of knowledge and skill in secondary education* (p. 39-52). Amsterdam: IDO/Vrije Universiteit.
White, T.G., Power, M.A., & White, S. (1989). Morphological analysis: implications for teaching and understanding vocabulary growth. *Reading Research Quarterly, 24*, 283-304.
White, T.G., Sowell, J., & Yanagihara, A. (1989). Teaching elementary students to use word-part clues. *The Reading Teacher, 42*, 302-308.
Wysocki, K. & Jenkins, J.R. (1987). Deriving word meanings through morphological generalization. *Reading Research Quarterly, 22*, 66-81.
Zechmeister, E.B., D'Anna, C.A., Hall, J.W., Paus, C.H., & Smith, J.A. (1993). Metacognitive and other knowledge about the mental lexicon: Do we know how many words we know? *Applied Linguistics, 14*, 188-206.

Manuscript ontvangen 16-5-1994
Definitieve versie ontvangen 22-11-1994

BIJLAGE 1

Voorbeelditems uit de toetsen Nederlands:

1. Metacognitietoets

Kun je de betekenis van het onderstreepte woord in de zin noemen?
Zet een cirkeltje om één van de cijfers.

	absoluut niet	waarschijnlijk niet	waarschijnlijk wel	zeker wel
En zo werd Jan <u>de dupe.</u>	1	2	3	4

2. Woordentoets

En zo werd Jan <u>de dupe.</u>

1 de laatste
2 de held van de dag
3 het slachtoffer
4 het mikpunt

3. Afleidtoets

Dat is een <u>flagrante</u> leugen. Je wordt knalrood en bovendien klopt je verhaal helemaal niet. Ze zagen je gisteren immers in de stad, terwijl jij zegt dat je de hele dag ziek in bed hebt gelegen.

flagrant = ...

4. Morfologietoets

<u>Pitten</u> zijn zaden die in vruchten zitten.
<u>Ontpitten</u> betekent .....

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 3, pp. 245-255

Aandachtsstrategieën bij het reviseren in een vreemde taal: de invloed van de taakomgeving

Hein Broekkamp[1] en Huub van den Bergh[2]
*Instituut voor de Lerarenopleiding, Universiteit van Amsterdam**
Vakgroep Taalbeheersing, Universiteit Utrecht

ABSTRACT

Revision processes put a demand on working memory. Given the relatively small capacity of working memory, an attention conflict may arise. It seems that especially foreign language writers are confronted with a conflict between attention for lower and higher level text demands. We hypothesize that the strategic choices necessitated by memory overload are guided by (a) the type and amount of errors in the text (b) explicit instructions with regard to the types of revisions required. In order to test this hypothesis an experiment was carried out. Dutch students (N=21) revised an English text in one of three conditions. In the first condition no extra problems or explicit instruction was added. The second condition offered the same task, except that 28 extra lower order problems were added. In the third condition, in addition to the extra problems, an explicit instruction was added to pay attention to higher order textual problems too. Results indicate that the addition of lower order problems diminishes the number of higher order revisions. As expected an explicit instruction seems to diminish the number of lower order revision. However the anticipated increase in the number of higher order textual problems could not be demonstrated. It is concluded that the coordination of attention is influenced by characteristics of the text to be revised. In L2-texts lower order characteristics seem to distract attention from higher order textual problems.

1. INLEIDING

Schrijven wordt wel beschreven in termen van jongleren met beperkingen (Flower & Hayes, 1980); schrijvers moeten rekening houden met allerlei restricties opgelegd door onder andere de opdracht, het publiek, de persoonlijke doelen, de tot dusver geproduceerde tekst, de taakomgeving en de individuele capaciteiten. Verder moeten ze op basis van deze restricties eisen afleiden voor het schrijfprodukt. Tenslotte moeten ze hun aandacht dienovereenkomstig verdelen. Al deze activiteiten maken deel uit van de coördinatie van aandacht voor verschillende teksteisen.

In deze bijdrage gaan wij in op de strategische aspecten van deze aandachtscoördinatie, en in het bijzonder op de coördinatie bij het reviseren in een vreemde- of tweede taal (hier beide aangeduid als T2). Wij richten ons op de vraag in hoeverre de taakomgeving het gebruik van aandachtsstrategieën benvloedt.

Om te komen tot een volwaardig schrijfprodukt dient de schrijver er zorg voor te dragen dat de regels van de spelling niet worden overtreden, maar ook dat de paragrafen op elkaar aansluiten, en dat de tekst aansluit bij de retorische situatie (het doel van de tekst, het publiek etc). Teksteisen variëren van de laagste niveaus (mechanische en linguïstische eigenschappen) tot hoogste niveaus (eigenschappen die nauw zijn gerelateerd met handelingsdoelen, zogenaamde pragmatische eigenschappen; Bereiter & Scardamalia, 1987, p.93; Kintsch & Van Dijk, 1983).

De schrijver moet de aandacht verdelen over verschillende teksteisen of -niveaus. Idealiter besteedt een schrijver aan al deze niveaus (in meer of mindere mate) aandacht. Gezien de beperkte capaciteit van het werkgeheugen brengt echter aanzienlijke aandacht op een bepaald moment voor het ene tekstniveau met zich mee dat aandacht voor andere tekstniveaus wordt

Adres: Herengracht 256, 1016 BV Amsterdam.

gereduceerd. Een dergelijke 'trade off' kan leiden tot een aandachtsconflict; problemen van het ene type worden niet opgemerkt of niet opgelost omdat de aandacht gericht is op problemen van een ander type (vergelijk: Stanovich (1980) voor een vergelijkbaar standpunt m.b.t. lezen). Aandachtsconflicten vormen dan ook een herkenbaar fenomeen. Zo citeren Bereiter & Scardamalia (1987) een zestienjarige schrijver: '*I have all my thoughts in my mind but when I come to a word I can't spell it trows me off my writing*' (p.154).

Glynn, Britton, Muth en Dogan (1982) laten zien dat aandachtsconflicten een aanzienlijke invloed kunnen hebben op de inhoudelijke kwaliteiten van het uiteindelijke schrijfprodukt. Zij lieten docenten een klad- en een netversie van een tekst schrijven en manipuleerden de aandacht voor teksteisen via een instructie. Het bleek dat hoe minder lagere orde teksteisen de docenten bij het schrijven van de eerste versie in beschouwing dienden te nemen, des te beter zij in staat waren aan *hogere orde* teksteisen bij de tweede versie te voldoen (het hoogste niveau betrof hierbij inhoudelijke kenmerken van de tekst). Verder stelden de onderzoekers vast dat een reductie van aandacht voor *lagere orde* teksteisen in de eerste versie *niet* ten koste ging van de tekstkwaliteit van de tweede versie op lagere niveaus. Hun verklaring voor deze bevindingen is dat bij de eerste versie, als gevolg van een reductie van teksteisen, meer geheugenruimte vrij kwam. Hieruit volgt dat schrijvers die in de eerste versie naast hogere ook lagere teksteisen in beschouwing moesten nemen, werden geconfronteerd met een overbelasting van het werkgeheugen, en de hieruit volgende aandachtsconflicten.

Daar meermalen is aangetoond dat het schrijven in een vreemde taal gekenmerkt wordt door een overmatige aandacht voor correct taalgebruik (Jones, 1985; Betancourt & Phinney, 1988), lijkt de kans op aandachtsconflicten tussen lagere en hogere orde teksteisen met name bij het schrijven in T2 toe te nemen. Wij veronderstellen dat een surplus aan aandacht voor linguïstische teksteisen ten koste gaat van de aandacht voor andere teksteisen en dat T2-schrijvers daardoor speciale moeilijkheden ondervinden om aan hogere orde teksteisen te voldoen. Deze veronderstelling wordt gesteund door een studie van Yau en Bereiter (besproken in: Bereiter & Scardamalia, 1987) naar het integreren van meer ideeën in een zin. Hieruit bleek dat Chinese studenten bij het schrijven in het Engels, hun tweede taal, minder ideeën integreerden dan hun medestudenten wier moedertaal Engels was.

De schrijver kan conflicten in aandacht voor verschillende teksteisen op verschillende manieren tegengaan. Bereiter en Scardamalia (1987, p.95) noemen er twee. Ten eerste kunnen aandachtsconflicten omzeild worden door automatisering; naarmate processen voor een groter deel onbewust plaats vinden – geen actieve aandacht vergen –, doen zij in geringere mate een beroep op het werkgeheugen. Ervaren schrijvers zijn, als gevolg van geautomatiseerde processen, grotendeels bevrijd van actieve aandacht voor eisen als correct taalgebruik. Onervaren schrijvers en T2-schrijvers zullen daarentegen in mindere mate beschikken over geautomatiseerde lagere orde processen dan respectievelijk ervaren schrijvers en T1-schrijvers. Een aandachtsconflict bij de eersten lijkt waarschijnlijker (zie ook: Bereiter & Scardamalia, p.176).

Ten tweede biedt de relatief grote tijdspanne van het schrijfproces, mogelijkheden om de aandacht zo te coördineren dat er geen of in mindere mate een 'trade off' plaats vindt in de aandachtsbesteding. Het al genoemde experiment van Glynn et al. laat zien dat voor het schrijven met behulp van twee versies een methode waarbij in de eerste versie alleen hogere orde teksteisen in beschouwing worden genomen, in mindere mate tot een aandachtsconflict leidt dan een methode waarbij in de eerste versie alle teksteisen in beschouwing worden genomen.

De aandachtscoördinatie kan sterk uiteenlopen tussen individuen. In een tweede experiment onderscheidden Glynn et al. (1982) hoog-verbale en laag-verbale studenten. Evenals bij de docenten lieten zij de twee groepen een tekst volgens twee versies schrijven, maar nu lieten zij de schrijvers vrij aan welke teksteisen zij bij de eerste versie aandacht besteedden (N.b. de teksteisen van de tweede versie waren wel vastgelegd in de instructie). Het bleek dat een aanzienlijk groter deel van de hoog-verbale groep dan de laag-verbale groep koos voor de effectieve decompositie-methode, waarbij in de eerste versie alleen aan hogere orde teksteisen aandacht werd besteed. In het eerste experiment van Glynn et al. was de decompositie van de

schrijftaak in versies en in aandachtsobjecten opgelegd, waardoor de taak gedeeltelijk van haar natuurlijke karakter ontdaan werd. In normale schrijfsituaties en tot op zekere hoogte ook in het tweede experiment van Glynn, bepaalt de schrijver zelf de aandachtscoördinatie.

Wij veronderstellen dat aandachtscordinatie in belangrijke mate wordt bepaald door aandachtsstrategieën. Waarbij strategieën verwijzen naar plannen die in een taaksituatie worden geconstrueerd aan de hand van informatie uit de taakomgeving en eigenschappen van het individu. Deze strategie fungeert dan als een richtlijn (controle-structuur) voor een relatief groot handelingspatroon van de taakuitvoering (Miller, Galanter & Pribram, 1960). In het geval van een aandachtsstrategie betreft het handelingspatroon de aandachtsbesteding en is de strategie zodoende een middel om met de beperkte capaciteit van het werkgeheugen om te gaan.

Aandachtsconflicten lijken ook bij het reviseren op te treden. Reviseren vereist immers ook in veel gevallen dat een omvattend geheel van teksteisen in beschouwing genomen dient te worden. In het onderzoek van Hayes, Flower, Schriver, Stratman & Carey (1985), komt in aandachtscoördinatie dan ook naar voren als cruciale component van het revisieproces. Voor deze coördinatie geldt dat de alom bekende decompositie-strategie ook bij het reviseren kan worden toegepast: de revisor besteedt eerst aandacht aan problemen van het ene type en daarna aan die van een ander type etc.: '*Editing could be done in several subtasks, each one attending to a specific aspect of the text*' (de Beaugrande, p. 215; en vgl het onderzoek van Glynn et al (1982), voor decompositie bij schrijven in het algemeen). Tot op zekere hoogte treedt decompositie bij reviseren vanzelf op. Bij elk probleem dat de revisor oplost, beperkt deze zijn aandacht tot één of een beperkt aantal teksteisen. Er kan daarnaast nog extremere decompositie plaatsvinden als de revisor gedurende langere tijd aandacht aan een teksteis besteedt, om vervolgens op een andere teksteis over te stappen. Dit betekent echter niet dat het revisieproces is bevrijd van aandachtsconflicten. Wij veronderstellen dat in het tijdelijk isoleren van de aandacht voor een teksteis, het gevaar schuilt dat overige teksteisen uit het oog worden verloren. Bovendien dient de schrijver bij het reviseren te ontdekken aan welke teksteisen aandacht moet worden besteed. Door de dominantie van één teksteis kan een schrijver volledig voorbij gaan aan andere teksteisen. Dergelijke coördinatieproblemen kunnen bijdragen aan een verklaring voor de bevindingen dat een groot aantal schrijvers bij het reviseren slechts in beperkte mate aandacht besteden aan hogere teksteisen (Fitzgerald, 1987).

In het theoretisch model dat Hayes et al. (1985) voor het revisieproces formuleren, wordt de aandachtscoördinatie uitgevoerd door het 'taakdefinitie'-proces: '*... task definition process reflects writer's conception of relevant revision criteria, revision activities and how to manage these activities over time ...*' (p.20). Taakdefinitie is dus het proces dat een representatie oplevert van de teksteisen en methoden om aan deze eisen te voldoen. Hayes et al zien taakdefinitie als een strategisch proces: '*In our description of revision then, the task definition process creates and stores 'metacognitive' or regulatory knowledge (...). As dynamic constructs, they depend partly upon stored conceptions (in long term memory) of what it means to revise and partly upon what information is provided (continually) by the task environment*' (p.21). De aandachtscoördinatie bij het reviseren wordt dus deels bepaald door de representatie van teksteisen. Wij veronderstellen dat deze representatie van teksteisen een grote invloed heeft op de uiteindelijke aandachtsbesteding en dat zij een strategisch karakter heeft. Dat wil zeggen: de revisor maakt gebruik van strategieën om haar aandacht tijdens het revisieproces te richten en de representatie van teksteisen maken deel uit van de strategieën, die op basis van de informatie uit de omgeving (bij voorbeeld: instructies) en de eigen kennis worden geconstrueerd.

Hoe krijgt de revisor een idee van de teksteisen waaraan de tekst (beter) dient te voldoen? Bij het vormen van een representatie van teksteisen kunnen schrijvers gebruik maken van uiteenlopende 'cues', informatie die naar de teksteisen verwijst. Deze 'cues' kunnen afkomstig zijn uit alle componenten van het schrijfproces, waaronder de taakomgeving (Hayes, 1992).

Richtlijnen uit de taakomgeving geven aan hoe gereviseerd dient te worden. Deze richtlijnen kunnen meer of minder expliciete aanwijzingen bevatten over het soort teksteisen waar de tekst

aan moet voldoen. Een expliciete aanwijzing is bij voorbeeld de richtlijn dat de betekenis van de tekst niet gewijzigd mag worden, een richtlijn die niet ongebruikelijk is bij het reviseren van gegeven teksten (Haugen, 1990). Vaak zullen richtlijnen echter impliciet zijn. De revisor wordt dan bij voorbeeld alleen gezegd 'dat de tekst verbeterd moet worden' of krijgt een beschrijving van de situatie te horen: een artikel moet bij voorbeeld verbeterd worden omdat het in eerste instantie niet geplaatst is in een tijdschrift. Bij het reviseren aan de hand van impliciete richtlijnen zal vanzelfsprekend een groter beroep worden gedaan op de actieve constructie van een representatie van teksteisen dan bij expliciete richtlijnen.

Andere 'cues' uit de taakomgeving zijn de gedetecteerde tekstproblemen. Het detecteren van tekstproblemen betekent voor de schrijver niet alleen dat deze het probleem moet oplossen, maar vormt tevens een signaal dat er problemen van betreffende type in de tekst kunnen zitten. Een schrijver die een aantal spelfouten heeft gedetecteerd, zal op zoek gaan naar andere spelfouten. Door de spelfouten wordt de aandacht gericht: de linguïstische teksteis neemt hierdoor een (dominantere) plaats in binnen de representatie van teksteisen.

Zoals hierboven gesuggereerd, veronderstellen wij dat bij het representeren van teksteisen een aandachtsconflict kan optreden. Als een teksteis een grote dominantie heeft verworven binnen de representatie van teksteisen, lijkt het moeilijk voor de schrijver om nog over te schakelen naar andere teksteisen. Een revisor die zich gedurende langere tijd met linguïstische problemen bezig heeft gehouden zal daardoor niet makkelijk over kunnen stappen op hogere orde problemen. Eén van de redenen voor dit dominantie-effect zou kunnen zijn dat het coördinatieproces taakdefinitie, evenals de aandachtsprocessen die het coördineert, een beroep doet op de capaciteit van het werkgeheugen. '*The executive competes for workspace with the subroutines it controls*' (Browne & Campione, 1981; geciteerd in Bereiter & Scardamalia, 1987, p. 111).

Een dergelijk aandachtsconflict, waarbij hogere orde aandacht gedomineerd wordt door lagere orde aandacht, lijkt met name op te treden bij T2-revisors. Zij veroorzaken, als gevolg van hun gebrekkige linguïstische vaardigheden, immers meer linguïstische tekstproblemen, en, gegeven hun preoccupatie met lingustische correctheid, zullen de problemen die zij detecteren, een sterk 'cue'-effect te hebben. Het lijkt derhalve interessant om te onderzoeken wat het effect is van een groot aantal linguïstische tekstproblemen bij het reviseren in T2. Uit het bovenstaande volgt dat als gevolg van een dominante 'cue'-werking een aandachtsconflict optreedt. Naarmate teksten van T2-revisors meer linguïstische problemen bevatten, zullen zij hogere orde teksteisen eerder uit het oog verliezen. Een dergelijk 'trade off'-effect wordt tegengegaan door expliciete informatie over teksteisen. Om deze twee veronderstellingen te toetsen is een experiment uitgevoerd dat hieronder wordt besproken.

2. METHODE VAN ONDERZOEK

In het experiment reviseerden 21 Nederlandstalige eerstejaars psychologie-studenten een gegeven Engelstalige tekst, die uiteenlopende problemen bevatte. De studenten zijn aselect aan één van de drie onderscheiden condities toegewezen. Deze drie condities verschilden in het aantal linguïstische tekstproblemen en de instructie die de aandacht van de studenten op bepaalde teksteisen moest richten.

Wij onderscheiden drie soorten teksteisen: linguïstische, tekstuele en pragmatische eisen. *Linguïstische teksteisen* verwijzen naar correct taalgebruik; de wijze waarop het taalsysteem als zodanig wordt gehanteerd dient correct te zijn. Dat wil zeggen: de spelling, grammatica en de letterlijke betekenis van woorden en zinnen dient juist te zijn. *Tekstuele teksteisen* verwijzen enerzijds naar abstracte maximes die te maken hebben met de informatie als zodanig en anderzijds naar maximes die betrekking hebben op de duidelijkheid waarmee de geboden informatie wordt verwoord. Het eerste soort maximes betekent dat de informatie volledig, coherent, juist en logisch moet zijn. Het tweede soort maximes betreft de eis dat de tekstinhoud eenvoudig, nauwkeurig en bondig is geformuleerd en helder is opgebouwd. Met *pragmatische teksteisen* tot

slot wordt verwezen naar de correspondentie tussen de tekst en de doelen van de lezer en de schrijver. Ook dienen deze doelen op zichzelf zinvol te zijn, dat wil zeggen: aan te sluiten bij de retorische situatie. Situationele factoren zijn: de relatie tussen de schrijver en het lezerspubliek (bij voorbeeld: formeel of persoonlijk), het onderwerp van de tekst (bij voorbeeld: reizen, of een sollicitatie), het teksttype (bij voorbeeld: beschouwend of informatief) en de setting (bij voorbeeld: studie of vrije tijd). Deze indeling naar teksteisen is gebaseerd op de verschillende betekenisniveaus van een tekst en sluit als zodanig aan bij Kintsch en Van Dijk (1983), en komt in grote lijnen overeen met de indeling van revisie-vormen van Hayes et al. (1985). Overeenkomstig Hayes et al. kan de tekstuele categorie opgevat worden als een restcategorie, waarbij noch formele linguïstische regels noch intenties van schrijver of lezer een rol spelen.

De revisietekst en de instructie zijn overgenomen uit het onderzoek van Hayes et al (1985). Deze tekst bevat substantiële aantallen linguïstische-, tekstuele en pragmatische problemen. Een belangrijk pragmatisch probleem is dat een brief van een sportlerares aan een collega omgezet moet worden in een A4-tje dat dient om meisjes te werven voor schoolsport. In de instructie wordt *impliciet* duidelijk gemaakt aan welke (pragmatische) eisen de tekst dient te voldoen: uit een situatiebeschrijving kan de revisor opmaken dat de tekst bestemd is voor schoolmeisjes en een wervend karakter dient te hebben.

De tekst is gemanipuleerd door 28 linguïstische problemen toe te voegen. Deze problemen betroffen grammaticale fouten (bij voorbeeld: *have learnt i.p.v. have learned*), spelfouten (*successfull i.p.v. successful*), woordvormingsfouten (*littler i.p.v. less*) en andere fouten (*coördination i.p.v. coordination*). In één van de condities zijn de teksteisen *expliciet* kenbaar gemaakt: de (impliciete) instructie is hier aangevuld met aanwijzingen welke teksteigenschappen voor verbetering in aanmerking kwamen:

- de meisjes moeten meer worden aangesproken en tot sporten aangezet worden;
- de structuur van de tekst moet verbeterd worden;
- de tekst moet begrijpelijker worden gemaakt;
- de spelling, de grammatica en de interpunctie dienen te worden verbeterd.

De eerste richtlijn betreft de pragmatische teksteisen, de tweede en derde richtlijn betreffen de tekstuele teksteisen, en de laatste richtlijn heeft betrekking op linguïstische teksteisen. In de andere conditie zijn de studenten niet expliciet gewezen op deze eisen. In de derde conditie tenslotte verbeterden de studenten een tekst zonder de toegevoegde problemen en zonder de aanvullende instructie. Op grond van de manipulatie van twee variabelen – aantal linguïstische fouten èn type instructie – zijn dus drie condities onderscheiden, die in tabel 1 zijn weergegeven.

Duidelijk blijkt dat verschillen tussen de conditie zonder toegevoegde fouten èn een impliciete instructie (ZI) en de conditie met toegevoegde fouten èn een expliciete instructie (ME) niet direct interpreteerbaar zijn in termen van één van de manipulaties. Een dergelijke interpretatie is alleen mogelijk door vergelijking van deze condities met de conditie met toegevoegde fouten èn een impliciete instructie (MI).

Tabel 1. Een overzicht van de manipulaties op grond waarvan drie condities onderscheiden worden.

Tekst	Instructie	Afkorting
Zonder toegevoegde linguïstische fouten	*I*mpliciet	ZI
*M*et toegevoegde linguïstische fouten	*I*mpliciet	MI
*M*et toegevoegde linguïstische fouten	*E*xpliciet	ME

Tijdens het reviseren, waarvoor onbeperkt de tijd is gegeven, zijn hard-op-denk-protocollen afgenomen, maar deze blijven hier buiten beschouwing. De taakuitvoering werd gevolgd door een 'cued recall'-interview, waarin de proefpersonen een toelichting hebben gegeven op de tekstwijzigingen die zij hadden aangebracht.

Aandacht voor linguïstische -, pragmatische -, en tekstuele teksteisen zijn geoperationaliseerd in termen van aantallen tekstwijzigingen. Uitgangspunt voor het indelen van tekstwijzingen vormde de soort eis die de revisor in beschouwing nam toen deze een verandering in de tekst aanbracht. Als indicaties dienden de tekstwijzingen zelf en de corresponderende opmerkingen uit het 'cued recall'-interview. Twee beoordelaars hebben de revisies geclassificeerd volgens de drie onderscheiden typen. De mate van overeenstemming tussen de twee observatoren die de problemen classificeerden, bleek redelijk (Cohen's Kappa = .66). De uiteindelijke code van discrepante observaties is gegeven na overleg tussen de twee beoordelaars. Opgemerkt zij dat met name de tekstuele categorie aanleiding gaf tot verwarring, wellicht doordat zij op verschillende wijze als restcategorie werd gehanteerd. Hierdoor kunnen resultaten die betrekking hebben op deze categorie slechts met voorbehoud worden geïnterpreteerd. De aandacht gaat daarom met name uit naar de resultaten die betrekking hebben op de linguïstische en pragmatische revisies.

Aan de hand van de eerst zin van de revisietekst worden hieronder de verschillende soorten tekstwijzingen toegelicht. Hierbij wordt ook het bijbehorende fragment uit het 'cued recall'-interview weergegeven, omdat de categorisering mede op basis hiervan plaatsvond.

(zin 1:) *Many naive girls students possess the assumption that it is necessary that they be superlative athletes in order to be successfull members of a school team.*

(1) Linguïstische tekstwijzingen: de revisor heeft zich zorgen gemaakt om correct taalgebruik.
naive -> naïve
wist ik niet ... dacht dat ik het zo moest schrijven;
possess -> posess
er moet een s weg ... ik weet niet welke.

(2 Tekstuele tekstwijzingen: de revisor heeft getracht de inhoud kloppend te maken en haar duidelijker te verwoorden (N.b. de revisor heeft zich noch zorgen gemaakt om formele linguïstische *regels*, noch om *intenties* van de schrijver of de lezer).
students possess -> students still possess
dat klopt sowieso niet;
possess the assumption -> assume
makkelijker lopen ... mooier lopen gewoon ... maakt eigenlijk niet zoveel uit.

(3) pragmatische tekstwijzingen: de revisor heeft geprobeerd de tekst beter te laten aansluiten bij de doelen van de lezer en de schrijver.
many naive girls -> many girls
ik vond dat stom van dat naief want dan scheld je die meisjes meteen uit ... je scheld ze niet uit maar beledigt ze eigenlijk meer;
possess the assumption that it is necessary -> think that it is necessary
ook te moeilijk ... stoppen ze gelijk met lezen ... het gaat over brugklassers ... meisjes volgens mij.

Tenslotte is de 'time on task' gemeten en is het 'totaal aantal wijzigingen' vastgesteld, om de mogelijke intermediërende invloed van deze algemene taakvariabelen vast te stellen. Naast een analyse van de geobserveerde aantallen wijzigingen lijken twee aanvullende analyses die rekening met deze beide variabelen houden wenselijk. Immers, naast verschillen in de verdeling van typen revisies kunnen verschillen in revisies tussen de condities zijn veroorzaakt door verschillen in het totale aantal gemaakte revisies of de tijd die aan de taak is besteed. Het is mogelijk dat

zowel de extra toegevoegde problemen als de aanvullende instructie, de studenten ertoe hebben aangezet meer tijd aan de taak te besteden of meer revisies te maken in het algemeen. Als gevolg hiervan zouden verschillen in de verdeling van type revisies – die juist interessant zijn als een indicatie voor de gehanteerde aandachtsstrategieën – verborgen kunnen blijven in analyses die het totale aantal revisies of de time on task niet in beschouwing nemen. Door (in statistisch opzicht) deze twee algemene taalvariabelen constant te houden is het mogelijk het effect van beiden op de verschillende typen revisies uit te sluiten.

Door toevoeging van linguïstische problemen wordt de aandacht op dit probleemtype gericht. Verwacht wordt dat dit leidt tot (verhoudingsgewijs) meer linguïstische revisies en minder pragmatische en tekstuele revisies. Door instructie wordt de aandacht (ook) op pragmatische en tekstuele teksteisen gericht. Hierdoor zal een revisor (verhoudingsgewijs) meer revisies aanbrengen die betrekking hebben op dergelijke teksteisen, bij een expliciete instructie dan bij een impliciete instructie.

3. RESULTATEN

In de onderstaande tabel zijn de gemiddelden aantallen revisies per type per conditie weergegeven.

Uit Tabel 2 blijkt dat het gemiddeld aantal revisies per categorie verschilt van conditie tot conditie. De verschillen suggereren effecten van de manipulaties, zoals verwacht. Toevoeging van linguïstische problemen in de te reviseren tekst lijken te leiden tot (logischerwijs) meer linguïstische revisies, maar ook tot minder pragmatische revisies: 3.5 (MI) vs. 8.3 (ZI). Hetzelfde geldt voor het effect van de aanvullende expliciete instructie. In de conditie waarin de proefpersonen expliciet op de teksteisen gewezen zijn, worden meer pragmatische revisies aangebracht: 3.5 (MI) vs. 7.3 (ZE). Conform de hypothese lijkt dit ten koste te gaan van het aantal linguïstische revisies. Echter, de relatief grote verschillen in zowel de tijd die aan de taak besteed is, als het totaal aantal fouten bemoeilijken een rechtstreekse interpretatie. Met name het tijdsverschil tussen de ZI- en MI-conditie is opmerkelijk. De toevoeging van linguïstische fouten lijkt aanleiding te geven tot een afname van de tijd die aan de taak besteed is.

Om te toetsen of de verschillen tussen de condities aan kans toe te schrijven zijn, zijn drie regressie-analyses uitgevoerd. In feite wordt een soort co-variantie-analyse uitgevoerd, waarmee het mogelijk is de invloed van bepaalde variabelen uit te partiëren. In de eerste analyse is dat niet gedaan: er is geen rekening gehouden met zowel time on task als het totaal aantal tekstwijzigingen. In de tweede analyse is daarentegen rekening gehouden met de verschillen in tijd die aan de opdracht besteed is. En, in de derde analyse tenslotte is rekening gehouden met het totaal aantal door een student aangebrachte tekstwijzigingen.

Tabel 2. Gemiddeld aantal gereviseerde fouten per type en per conditie. (ZI: zonder toegevoegde linguïstische problemen èn impliciete instructie; MI: met toegevoegde linguïstische èn impliciete instructie; ME: met toegevoegde linguïstische èn expliciete instructie; $\overline{X}$: gemiddelde; σ: standaarddeviatie; N: aantal proefpersonen).

	Linguïstisch		Pragmatisch		Tekstueel		Tijd		
Conditie	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ	N
ZI	5.8	6.0	8.3	9.2	7.4	5.4	42.9	27.9	7
MI	28.0	12.4	3.5	5.0	7.7	2.8	37.1	7.2	7
ME	20.8	6.8	7.3	6.9	9.8	4.3	55.1	21.1	7

Tabel 3. Verschillen tussen condities per analyse per revisie-categorie. (ZI: zonder toegevoegde linguïstische problemen èn impliciete instructie; MI: met toegevoegde linguïstische èn impliciete instructie; ME: met toegevoegde linguïstische èn expliciete instructie; Analyse I: absolute verschillen; Analyse II: conditioneel op 'time on task'; Analyse III: conditioneel op totaal aantal revisies; r^2: proportie verklaarde variantie; standaardfouten tussen haakjes).

Cond.	Teksteis	Analyse I	Analyse II	Analyse III
ZI	Linguïstisch-	−22.2 (4.4)	−23.0 (4.6)	−16.0 (3.6)
	Pragmatisch	4.8 (3.6)	4.7 (2.0)	11.0 (3.3)
	Tekstueel	−0.3 (2.1)	−1.1 (2.0)	5.9 (1.9)
MI	Linguïstisch	27.9 (3.1)	29.1 (3.3)	25.8 (2.8)
	Pragmatisch	3.5 (2.5)	4.6 (1.9)	1.3 (2.3)
	Tekstueel	7.7 (1.5)	8.8 (1.4)	5.5 (1.3)
ME	Linguïstisch	−7.0 (3.4)	−9.5 (4.7)	−6.6 (3.5)
	Pragmatisch	3.8 (3.6)	1.3 (2.8)	4.2 (3.2)
	Tekstueel	2.1 (2.1)	−0.3 (2.0)	2.5 (1.8)
	Tijd		0.1 (.03)	
	Totaal			0.4 (.02)
	r^2	.58	.62	.70

In al de analyses zijn de verschillen tussen de drie condities geschat[1]. Om de interpretatie van de resultaten te vergemakkelijken zijn de gemiddelden voor de conditie zonder toegevoegde fouten èn een impliciete instructie (ZI) èn de conditie met toegevoegde fouten èn een expliciete instructie (ME) geschat als afwijkingen van de conditie met toegevoegde fouten en een impliciete instructie (MI). Aan de hand van de ratio van de schattingen en de bijbehorende standaardfouten kan direct bepaald worden of de gemiddelden voor één van de probleemtypen in de condities ZI en ME significant afwijkt van die in de conditie MI (|schatting| / standaardfout > 1.96; $p < .05$).

Op grond van de parameterschattingen in het eerste regressiemodel (analyse I) kunnen eenvoudig de cel-gemiddelden voor de drie foutsoorten per conditie bepaald worden. In de conditie met extra problemen en een impliciete instructie zijn de celgemiddelden rechtstreeks geschat. Vergelijking van deze parameterschattingen met de geobserveerde gemiddelden (Tabel 2) leert dat de gemiddelden alleszins redelijk (terug)voorspeld worden.

Om de gemiddelden te berekenen in de conditie met extra problemen èn een expliciete instructie (ME), dan wel de gemiddelden in de conditie zonder extra problemen en een impliciete instructie (MI), moeten we bedenken dat deze geschat zijn als afwijking van die van de conditie met extra problemen en een impliciete instructie. Dat wil zeggen: de drie cel-gemiddelden voor de ME- conditie zijn geschat als (3.5 + 3.8 =) 7.3 voor pragmatische, (27.9 − 7.0 =) 20.9 voor de linguïstische problemen en als (7.7 + 2.1 =) 9.8 voor de tekstuele-categorie. Voor de MI-conditie zijn de celgemiddelden respectievelijk als volgt geschat (3.5 + 4.8 =) 8.3, (27.9 − 22.2 =) 5.7, en (7.7 − .3 =) 7.4.

Op grond van de ratio van de (absolute waarde van de) parameterschattingen en de standaardfouten kan nu direct bepaald worden of de verschillen ten gevolge van de toevoeging van linguïstische problemen, dan wel de wijziging van de instructie significant zijn. Zo zijn volgens analyse I verschillen tussen de ZI en MI-conditie in pragmatische revisies niet significant: 4.8 / 3.6 < 1.96. Voor de linguïstische revisies is dat duidelijk wel het geval:
|-22.2 / 4.4| > 1.96.

In de tweede en derde analyses zijn de celgemiddelden niet zo eenvoudig terug te berekenen

als in analyse I, daar beide genoemde variabelen een duidelijk effect hebben op het aantal revisies in de onderscheiden categorieën.

Overigens konden geen verschillen in regressiegewichten worden aangetoond voor zowel het totaal aantal revisies als de tijdsbesteding. Omdat de regressie van deze beide covariaten constant is over condities kan volstaan met één te schatten parameter voor de drie condities tezamen.

Uit alle analyses blijkt dat, zoals voorspeld, significant minder linguïstische problemen worden gereviseerd in de ZI-conditie dan in de MI-conditie. Opvallender is dat ook in de ME conditie minder linguïstische problemen worden gereviseerd. Dit geldt echter niet voor de derde analyse. Voor de ander typen problemen kunnen met analyse 1 geen verschillen worden vastgesteld. Analyse 2 en 3 brengen echter wel verschillen aan het licht tussen de ZI- en de MI-conditie: in de eerstgenoemde wordt een groter aantal pragmatische revisies gemaakt. Dergelijke verschillen zijn er niet tussen de ME en MI-conditie.

Samenvattend: conform de in de inleiding genoemde hypothese kunnen tussen de beide condities met een impliciete instructie (ZI en MI) verschillen in aantallen revisies per categorie aangetoond worden. Deze verschillen ten gevolge van de manipulatie van het aantal linguïstische fouten blijken het duidelijkst als in de analyse rekening gehouden wordt met het totaal aantal aangebrachte revisies. Maar ook als rekening gehouden wordt met de tijd die aan de taak besteed is komt een verschil in aantal pragmatische revisies als gevolg van de genoemde manipulatie naar voren.

De verschillen ten gevolge van de manipulatie van de instructie zijn minder duidelijk; alleen voor het aantal linguïstische revisies kan in twee van de drie analyses (de geobserveerde aantallen en conditioneel op tijd besteed aan de taak) een verschil geïnduceerd door de instructie aangetoond worden.

4. DISCUSSIE

In deze studie zijn T2-revisiestrategieën van eerstejaars studenten onderzocht. Getracht is de revisiestrategie te beïnvloeden via 'cues' uit de taakomgeving, namelijk: toevoegingen van linguïstische problemen in de te reviseren tekst en expliciete informatie over teksteisen. Het is gebleken dat toevoeging van de linguïstische problemen het verwachte effect had: het leidde tot relatief meer linguïstische en pragmatische revisies, wanneer rekening wordt gehouden met het totaal aantal gemaakte revisies of de tijd die aan de taak is besteed. Voor de expliciete instructie kon echter geen significant effect worden vastgesteld.

Uit het bovenstaande kan worden geconcludeerd dat extra linguïstische problemen tot een andere aandachtsverdeling leiden. Het feit dat geen absolute verschillen in pragmatische revisies konden worden aangetoond (analyse I), suggereert dat verminderde aandacht voor een teksteis op een bepaald moment, wordt gecompenseerd door de revisors doordat zij meer tijd aan de taak besteden of doordat zij meer revisies maken. Echter, dit experiment met zijn relatief beperkte onderscheidingsvermogen leent zich niet voor ferme conclusies.

Voor het uitblijven van een significant effect van instructie zou ook een statische verklaring – in termen van onderscheidend vermogen – kunnen worden gegeven. Niettemin, suggereren de resultaten een verschil in effectiviteit tussen deze manipulatie en de toevoeging van linguïstische problemen. Dit verschil kan als volgt theoretisch geduid worden. Als revisors hun aandachtsbesteding tot stand laten komen door de problemen die zij in de tekst tegenkomen, passen zij hun aandachtsstrategie voortdurend aan op grond van nieuwe 'cues' uit de taakomgeving. Een dergelijke revisie-aanpak kan opportunistisch genoemd worden (vergelijk: Hayes-Roth & Hayes-Roth, 1979). Tegengesteld aan zo'n opportunistische aanpak is een methode waarbij revisors aandachtsstrategieën baseren op informatie die relatief stabiel aanwezig is tijdens de taakuitvoering. Dit is bij voorbeeld het geval als instructie het strategiegebruik beïnvloedt. In deze studie lijkt het dat een opportunistische methode die gericht was op linguïstische teksteisen

dominant was over de andere methode, wellicht doordat zij vaker werd 'bekrachtigd' door nieuwe linguïstische detecties. In dit verband is het aardig op te merken dat uit de hard-op-denk-protocollen blijkt dat naarmate de instructie vaker doorgelezen wordt, dan wel uit het lange duur geheugen opgediept wordt, ook meer pragmatische revisies plaatsvinden. Achteraf gezien is het jammer dat er niet nog een vierde conditie is onderscheiden met een expliciete instructie, maar zonder toegevoegde linguïstische problemen. Het was dan mogelijk geweest om de afzonderlijke bijdrage van instructie vast te stellen.

Uit het experiment zoals het is uitgevoerd, dienen wij ons te beperken tot de conclusie dat bij het reviseren in een vreemde taal, linguïstische tekstproblemen een dominante invloed uitoefenen, die niet makkelijk lijkt te kunnen worden tegengegaan door een korte instructie. Deze conclusie zien wij als bevestiging van drie algemenere noties, waaruit weer onderwijsimplicaties volgens. Het eerste idee is dat de taak een aanzienlijke invloed uitoefent. Een implicatie hiervan voor het schrijfonderwijs is dat schrijftaken afgestemd dienen te worden op leerdoelen. Als het doel bijvoorbeeld is hogere orde kenmerken van een tekst te verbeteren, moeten taken die een grote hoeveelheid lagere orde aandacht vereisen worden vermeden. Het tweede idee betreft de coördinatie van het schrijfproces. Het idee is dat de beperkingen (eisen) waarmee de schrijver jongleert niet altijd gelijk van gewicht zijn en dat zwaarwegende beperkingen ervoor kunnen zorgen dat de schrijver andere beperkingen laat vallen. Dit idee impliceert dat, naast schrijfvaardigheden die gericht zijn op afzonderlijke taakeisen, schrijvers over strategieën dienen te beschikken om hun aandacht voor deze eisen te coördineren. Het derde idee, tenslotte is verweven met de andere twee. Het idee is dat aandachtsconflicten tussen lagere en hogere tekstniveaus eerder manifest worden voor T2-schrijvers dan voor schrijvers in hun moedertaal. Zowel taakverschillen – voor revisietaken lijkt de dichtheid van linguïstische problemen in hun teksten hoger – als een preoccupatie met correct taalgebruik, zorgen ervoor dat T2-schrijvers jongleren met een extra zware beperking van de taal. Een implicatie is dat T2-schrijvers uitgerust dienen te worden met effectieve aandachtsstrategieën, tenzij de eisen van hun schrijfprodukten gewijzigd wordt.

NOOT

1. In feite zijn de gepresenteerde (regressie-)analyses identiek aan (co)variantie-analyses, met dien verstande dat er bij de regressie-analyses drie residuele varianties, voor elke teksteis één, geschat kunnen worden. Hierdoor kunnen de gegevens van de drie teksteisen en de drie condities simultaan geanalyseerd worden.

LITERATUUR

Bereiter, C. & Scardamalia M. (1987). *The psychology of written composition*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Betancourt, F. & Phinney, P. (1988). Sources of writing block in bilingual writers. *Written communication, 5*, 461-478.

Browne, A.I. & Campione, J.C. (1981). Inducing flexible thinking: a problem of access. In M. Friedman, J.P. Das & N. O'Connor (eds.). *Intelligence and learning* (515-529). New York: Plenum.

Fitzgerald, F. (1987). Research on revision and writing. *Review of Educational Research, 57*, 481-506.

Flower, L. & Hayes, J.R. (1980). The dynamics of composing: making plans and juggling constraints. In L.W. Gregg & E.R. Steinberg (eds.). *Cognitive processes in writing*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Glynn, S.M., Britton, B.K., Muth D.K. & Dogan, N. (1982). Writing and revising persuasive documents: cognitive demands. *Journal of Educational Psychology, 74*, 557-567.

Haugen, D. (1990). Comming to terms with editing. *Research in the Teaching of English, 24*, 323-333.

Hayes, J.R. (1992). Naar een nieuw model voor tekstproduktie: een voortgangsverslag. *Tijdschrift voor Taalbeheersing, 14*, 247-256.

Hayes, J.R., Flower, L., Schriver, L., Stratman, J. & Carey, L. (1985). *Cognitive processes in revision* (Technical report). ERIC document, ED 267 396.

Hayes-Roth B. & Hayes-Roth, F. (1979). A cognitive model for planning. *Cognitive Science, 3*, 275-310.

Jones, S. (1985). Problems with monitor use in second language composing. In M. Rose (ed.). *Studies in writers' block and other composing process problems* (90-118). New York: Guilford.

Kintsch, W. & Dijk, T. van (1983). *Strategies in discourse comprehension*. London: Academic Press.

Miller, G.A., Galanter, E. & Pribram, K.H.(1960). *Plans and the structure of behaviour*. New York: Holt, Rinehart & Winston.

Stanovich, K.E. (1980). Toward in interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quartely*, 16/17, 32-71.

Manuscript ontvangen 16-5-1994
Definitieve versie ontvangen 22-11-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 3, pp. 256-270

Leesstrategieën van anderstalige en Nederlandse leerlingen in het voortgezet onderwijs

Hilde Hacquebord
Vakgroep Taal en Communicatie, Faculteit der Letteren, Rijksuniversiteit Groningen

ABSTRACT

In this article the results of two research projects are reported. The first project questions the differences between L1 and L2 readers with respect to the linguistic variables Word Knowledge, and Grammatical Understanding in Dutch as a Second Language, and to the nonverbal variables Knowledge of Text Subject and Non-verbal IQ in relationship to Text Comprehension. The L2 readers are deficient on the linguistic variables only, which was expected. With respect to Text Comprehension, they seem to compensate for their lack of word and grammatical knowledge by applying top down strategies in reading for comprehension. This was explored by analyzing their reading behaviour in so-called score-profiles in which three aspects of Text Comprehension were distinguished: micro-, meso- and macro-comprehension as different levels of textual processing. The L2-readers appeared to perform the worst at the micro-level, both in comparison with the Dutch readers and in comparison with their results on the macro-subtest. The conclusion was drawn that L2-readers compensate for their problem in word-processing by a relatively strong emphasis on the comprehension of the macro-level, which indicates a top-down processing style.
In a follow-up project weak readers in general were involved, among them L2-readers but also dyslectic readers. In this project the idea was explored to improve text comprehension of weak readers by making use of their capacity to use top-down strategies in a compensating way. A course was designed in which top-down reading strategies were taken as the starting point for a reading programme. We hoped to find positive results for all types of readers involved in the program, which was indeed the case with respect to general text comprehension. With respect to specific aspects of text comprehension, only the Dutch pupils improved their scores. There is a need to improve the reading programme, especially to meet the problems of L2-readers with respect to the micro- and mesoprocessing of texts.

1. TAAL, KENNIS EN LEESVAARDIGHEID

Tot voor kort werd lezen beschouwd als een vaardigheid die leerlingen in het basisonderwijs voldoende onder de knie krijgen om het voortgezet onderwijs te kunnen doorlopen. Inmiddels is duidelijk dat Nederlandse leerlingen met name op basisschoolleeftijd in internationaal opzicht relatief slecht presteren als het gaat om leesbegrip (De Glopper & Otter, 1993a). In het basisonderwijs zou de aandacht wel eens tezeer op het technische aspect van lezen gericht kunnen zijn; meer aandacht voor begrijpend lezen in het basisonderwijs lijkt noodzakelijk (De Glopper & Otter, 1993b).

In het voortgezet onderwijs moeten leerlingen kennis verwerven uit teksten; in het vakonderwijs wordt een groot beroep gedaan op hun leesbegrip, met name van schoolboekteksten. Met name anderstalige leerlingen blijken problemen te hebben met de leestaken die in het voortgezet onderwijs worden gesteld, waarvan het lezen en verwerken van schoolboekentaal een van de weinige onderzochte thema's is (Hofmans-Okkes, 1987; Hacquebord, 1989; De Jong, 1990). Analyses van schoolboekteksten maken duidelijk dat hierin heel wat taalkennis en taalvaardigheid van leerlingen vereist wordt: ze bevatten veel voor leerlingen onbekende of moeilijke woorden, complexe grammaticale constructies en ondoorzichtige tekstuele structuren. Voordat we het (tweede-)taalonderwijs echter gaan richten op het aanleren van dergelijke typerende taal-

Adres: Oude Kijk in 't Jatstraat 26, 9712 EK Groningen.

en tekstverschijnselen – hoe, dat is dan nog een andere vraag – is een nadere analyse van het 'tekstbegripprobleem' op zijn plaats. Bij tekstbegrip spelen namelijk, behalve talige variabelen, ook non-verbale variabelen een belangrijke rol.

Niet alleen anderstaligen hebben problemen met het begrijpen van schoolboekteksten. Ook veel Nederlandse leerlingen weten geen raad met de taak zelfstandig te leren uit teksten, die ook voor Nederlandssprekende leerlingen veel moeilijke woorden en constructies bevatten. In het probleemloze geval leert een leerling door vallen en opstaan om te gaan met schoolboekentaal. Succesvolle leerlingen ontwikkelen uit zichzelf en vaak onbewust strategieën waarmee ze teksten aanpakken en begripsproblemen oplossen. Maar lang niet alle leerlingen komen tot efficiënt strategiegebruik bij tekstbegrip (Brown, 1985). Problemen met het verwerken van teksten worden in het geval van moedertaallezers ('T1-lezers') niet in de eerste plaats toegeschreven aan taalproblemen, maar eerder aan deficiënties op het gebied van voorkennis en strategische competentie.

Het begrijpend kunnen lezen van teksten is een complex proces waarbij interacterende factoren een rol spelen en dat bij verschillende lezers op verschillende manieren tot stand kan komen, ook afhankelijk van de leestaak en het begripscriterium dat een lezer zichzelf stelt (Van Oostendorp, 1988). Vanuit het perspectief van de lezer gezien, is het van belang dat een tekst voldoende houvast biedt voor adequaat begrip. Een tekst is voor de lezer moeilijker, naarmate er meer onbekende zaken in voorkomen. Ter illustratie: de ervaren lezer die geen Italiaans kent, kan toch veel te weten komen uit een Italiaanse krant. De begrijpelijkheid van dergelijke teksten in een vreemde taal hangt vooral af van de mate waarin elementen uit die teksten bekend zijn: inhoudelijke zaken zowel als taal- en tekstkenmerken. Daarbij speelt strategische competentie een rol: in hoeverre weet een lezer gebruik te maken van kennis in een poging te komen tot een betekenisreconstructie. Die betekenisreconstructie van de lezer is dus het resultaat van een samenspel tussen taalvaardigheid, kennis van het tekstonderwerp en strategie. Bernhardt (1991) onderscheidt in dit verband drie clusters van interacterende variabelen:

1. Linguïstische variabelen, die de manifeste, verbale elementen in een tekst vormen, zoals woordstructuur, woordbetekenis, syntaxis and morfologie;
2. Leesvaardigheidsvariabelen, waaronder leesstrategieën en
3. Kennisvariabelen, betreffende de nonverbale (achtergrond)kennis van de lezer met betrekking tot het tekstonderwerp.

Onderzoek bij tweede-taallezers ('T2-lezers') heeft duidelijk gemaakt dat leesproblemen primair worden veroorzaakt door taaltekorten, met name een gebrek aan woordkennis dat als specifiek T2-probleem in onderzoek telkens weer naar voren komt (Alderson, 1984; Bossers, 1992). Aan de andere kant zijn er aanwijzingen gevonden dat anderstaligen juist door hun taalproblemen een relatief sterker beroep doen op hun strategische vaardigheid. In onderzoek van Ulijn en Strother (1990) behaalden T2-lezers zelfs betere tekstbegripprestaties dan de T1-lezers met eenzelfde kennisniveau. De onderzoekers verklaren dit verschijnsel uit het verschil in attentiewaardie die een tekst voor lezers met verschillende taalachtergronden heeft: van de anderstalige lezers vraagt een (inhoudelijk interessante) tekst meer inspanning en aandacht dan van de T1-lezers. De T1-lezers bleken de tekst veel sneller te lezen dan de T2-lezers, die waarschijnlijk een diepere verwerking bereiken. Uit het onderzoek van Jetten (1991) blijkt dat T2-lezers meer dan T1-lezers een top-down manier van lezen demonstreren bij tekstfragmenten waarin bekende 'scripts' worden beschreven, dat wil zeggen stereotype gebeurtenissen die verwijzen naar de cultuurbepaalde kennis van de wereld (bijvoorbeeld 'eten in een restaurant'). Doordat de integratie van informatie in een scripttekst gemakkelijker is, bereiken T2-lezers met name een grotere mate van woordherkenning dan in teksten zonder bekend script. Ook lezers die vanwege andere problemen moeite hebben met het verwerken van de linguïstische input, zoals bijvoorbeeld dyslectici, blijken dit tekort te compenseren door een grotere inzet van strategische vaardigheden en/of door meer gebruik te maken van contextkennis (Stanovich, 1980).

De lezers uit bovengenoemd onderzoek zijn op een bepaalde manier 'gehandicapt' en juist

daardoor zien we dat ze (compenserende) strategische vaardigheden ontwikkelen. Zij maken van de nood een deugd. 'Gewone' leerlingen zullen uit zichzelf veel minder snel tot een strategisch leesgedrag komen, hoewel dit evenzeer gewenst is. Voor alle leerlingen geldt immers dat ze in staat moeten zijn zich 'in nood' te redden, dus bijvoorbeeld bij nieuwe, onbekende (lees)taken zoals die in het vakonderwijs worden gesteld. In de Basisvorming zou daarom aan alle leerlingen onderwijs in functionele leesvaardigheid moeten worden geboden waarbij de aanpak, dus ook het toepassen van leesstrategieën, centraal staat.

In het hiervolgende zal worden gerapporteerd over twee projecten. Ten eerste wordt een aantal onderzoeksresultaten gepresenteerd die in Hacquebord (1989), vanuit een longitudinale vraagstelling, uitgebreider worden gerapporteerd. Hieruit kunnen we afleiden dat T1- en T2-leerlingen in verschillende schooltypen (lbo en mavo) bij functioneel tekstbegrip verschillende leesstrategieën hanteren, die deels afhankelijk zijn van hun kennis en taalvaardigheid. Het verschil in het bereikte niveau van tekstbegrip tussen Nederlandstaligen en anderstaligen is overigens minder groot dan het verschil tussen leerlingen van verschillende schooltypen. Met name lbo-leerlingen blijken in het algemeen zwakke lezers te zijn.

In een vervolgproject op dit dissertatieonderzoek wordt het optreden van eventuele verschillen in leesstrategieëen bij verschillende categorieën leerlingen nader geëxploreerd. Op basis van diagnostisch onderzoek zijn op een aantal scholen voor voortgezet onderwijs 'leeszwakke' leerlingen geselecteerd, onder wie nogal wat anderstaligen, maar ook leerlingen met leerachterstanden en -stoornissen, waaronder een aantal dyslectische leerlingen. Deze leerlingen hebben een experimentele leescursus gevolgd, gericht op een verbetering van hun tekstbegrip, waarbij de nadruk werd gelegd op het bevorderen van strategisch leesgedrag. Over de effecten van dit interventieprogramma wordt in de laatste paragraaf gerapporteerd.

2. TEKSTBEGRIP VAN TURKSE EN NEDERLANDSE LEERLINGEN IN HET VOORTGEZET ONDERWIJS

Het onderzoeksproject naar tekstbegrip van T1- en T2-leerlingen in het voortgezet onderwijs (Hacquebord, 1989) heeft een longitudinale opzet, waarbij leerlingen vanaf de brugklas tot in de derde klas van het voortgezet onderwijs zijn gevolgd. In het eerste jaar zijn behalve tekstbegrip ook woordenschat, grammatica, non-verbaal IQ en voorkennis van het tekstonderwerp getoetst. Gedurende de onderzoeksperiode is tekstbegrip bij dezelfde leerlingen elk jaar getoetst met geparallelliseerde tekstbegriptoetsen. In het hiervolgende zal niet worden ingegaan op de longitudinale resultaten die zicht geven op de ontwikkeling van het tekstbegrip van leerlingen in de eerste jaren van het voortgezet onderwijs. Hier wordt volstaan met de samenvatting van deze longitudinale resultaten: er was geen sprake van duidelijke vooruitgang in tekstbegrip, bij geen van de groepen leerlingen die zijn gevolgd. Eerder was er sprake van een zgn. Mattheuseffect (Stanovich, 1986): de betere lezers (in dit geval Nederlandse mavo-leerlingen) werden wat beter, de zwakke lezers (Turkse lbo-leerlingen) slechter.

De algemene vraag die hier centraal staat is of de onderzoeksresultaten bovendien aanwijzingen kunnen opleveren voor eventuele verschillen in het hanteren van leesstrategieën door de verschillende groepen leerlingen. Door een aantoonbaar tekort op het gebied van taal- en woordkennis zouden anderstalige leerlingen immers geneigd kunnen zijn bij tekstbegrip sterker te leunen op non-verbale kennis en strategische vaardigheden.

De onderzoeksvragen die in dit verband aan de orde worden gesteld zijn:

1. Wat zijn de verschillen tussen Nederlandse (T1) en Turkse (T2) lbo en mavo-leerlingen aan het begin van het voortgezet onderwijs met betrekking tot (a) linguïstische aspecten (woordkennis en grammaticaal begrip van het Nederlands), (b) non-verbale aspecten (voorkennis en non-verbaal IQ) in relatie tot (c) het begrip van teksten, afkomstig uit schoolboeken?
2. Zijn er verschillen tussen T1 en T2-leerlingen als het gaat om de aard van hun tekstbegrip? Kunnen we verschillen in leesstrategieën afleiden uit verschillen in scoreprofielen?

2.1 Onderzoeksopzet

Uit brugklassen van zeven verschillende scholen voor voortgezet onderwijs in Hengelo en Enschede zijn in 1985/86 in principe alle onderinstromende Turkse leerlingen in het onderzoek betrokken (n= 88). Deze leerlingen stromen allen door naar mavo of lbo. Een vrijwel even grote groep van Nederlandse leerlingen (n=89) is geselecteerd op basis van zoveel mogelijk dezelfde kenmerken als die van de Turkse groep: de samensteling van de Nederlandse groep is vrijwel gelijk als het gaat om sekse, leeftijd, school, schooltype, klas. De groepen verschillen in talige en (sociaal-)culturele achtergrond.

Bij deze groepen zijn in het eerste brugjaar toetsen afgenomen voor de in vraagstelling 1 genoemde variabelen. De toetsen voor grammatica, woordenschat en voorkennis betreffen driekeuze-toetsen en bevatten stof die representatief kan worden geacht voor het schriftelijke schooltaalaanbod in de brugklas (Galema, 1988). De toets voor tekstbegrip is eveneens gebaseerd op een analyse van het schriftelijke taalaanbod in het voortgezet onderwijs. Deze toets bestaat uit 72 items, gelijk verdeeld over zes representatieve teksten uit schoolboeken op drie moeilijkheidsniveaus. De non-verbale toets is de bekende Raven Standard Progressive Matrices (Raven, Court en Raven, 1977). Door middel van tweeweg-variantie-analyse worden de resultaten van Turkse en Nederlandse lbo- en mavoleerlingen geanalyseerd en getoetst.

Bij de analyse die gemaakt wordt voor het beantwoorden van vraagstelling 2 wordt de tekstbegriptoetsscore geanalyseerd in drie deelscores van elk 24 items voor drie te onderscheiden aspecten van tekstbegrip: micro- meso en macrobegrip. Door multivariate variantie-analyse met deze drie subscores als herhaalde metingen wordt een profielanalyse gemaakt die mogelijk inzicht biedt in de aard van het tekstbegrip, dat bij verschillende (groepen) leerlingen op verschillende manieren tot stand kan zijn gekomen, bijvoorbeeld doordat ze verschillende leesstrategieën hanteren. De verwachting is dat Nederlandse leerlingen sterker dan de Turkse leerlingen bij tekstbegrip hun taalkennis kunnen aanwenden en dat dit blijkt uit een relatief hogere score op micro- en mesosubtoetsen, terwijl de Turkse leerlingen ter compensatie van hun tekort aan taalkennis een relatief grotere macrotekstbegripscore zullen laten zien. In verband met deze veronderstelling over verschillen in leesstrategieën, die worden afgeleid van verschillen in scores op onderscheiden subtoetsen voor zogenaamd micro- meso- en macrotekstbegrip, zal hier eerst nader worden ingegaan op de constructie van de tekstbegriptoets.

2.1.1 De constructie van de tekstbegriptoets

De vraag of er verschillen zijn in het aan tekstbegrip ten grondslag liggende leesgedrag van de T1 en T2-leerlingen (vraagstelling 2) proberen we te beantwoorden door de (sub)scores op de tekstbegriptoets verder te analyseren. Evidentie voor een top-down strategische leesstijl bij de Turkse leerlingen kan dan blijken uit een analyse van de scoreprofielen die worden opgesteld op basis van de micro- meso- en macro-subscores van de tekstbegriptoets. De manier van tekstbegriptoetsing die hierboven ter sprake kwam, behoeft hier enige toelichting omdat hierna de termen 'micro-', 'meso'- en 'macro'-leesvaardigheid worden gebruikt bij het opstellen van verschillende lezersprofielen. Deze drie vormen van leesvaardigheid worden hier beschouwd als aspecten van tekstbegrip die in verschillende mate door leerlingen worden beheerst en/of toegepast als ze lezen. De toets is geconstrueerd op basis van een analyse van de tekststructuur waarin we een micro- een meso- en een macroniveau van de tekst onderscheiden. Een dergelijke analyse is ontleend aan de tekstlinguïstiek van o.a. Van Dijk en Kintsch (1983). Ook de vraagvorm is geïnspireerd door de tekstlinguïstiek waarin de dieptestructuur van teksten wordt blootgelegd in een hiërarchische set van proposities, die we in dit verband moeten opvatten als enkelvoudige tekstuitspraken op verschillende tekstuele niveaus.

Een macropropositie heeft betrekking op de tekst in zijn geheel, het hoofdthema, de functie, de 'overall' tekststructuur. Een mesopropositie betreft een verbinding die in de tekst wordt gelegd tussen zinnen binnen een alinea. Een micropropositie is een uitspraak op het niveau van woorden en zinnen. De proposities worden in de toets als stellingen worden geformuleerd, waarop een leerlingen met 'klopt' of 'klopt niet' moeten reageren (zie appendix voor voorbeeld). De tekstbegriptoets bevat zes teksten met elk 12 vragen, waarvan telkens vier micro-,

vier meso- en vier macrostellingen betreffen. De toets heeft dus in totaal 72 items (toetsbetrouwbaarheid a= .80), en de deeltoets Micro (toetsbetrouwbaarheid a= .61), deeltoets Meso (toetsbetrouwbaarheid a= .54) en deeltoets Macro (a=.52) bevatten elk 24 items. De relatief geringe betrouwbaarheid van de deeltoetsen wordt veroorzaakt door het kleine aantal items. In de longitudinale analyse wordt dit bezwaar ondervangen door herhaalde metingen met parallelle subtoetsen. Het feit dat de drie deeltoetsen geconstrueerd zijn aan de hand van hetzelfde tekstmateriaal en daardoor bij voorbaat een sterke onderlinge samenhang vertonen, kan eveneens als een zwak punt worden beschouwd, dat evenwel bij de gekozen opzet onvermijdelijk is. In plaats van over afzonderlijke tekstbegripdeelvaardigheden, is het daarom beter te spreken over te onderscheiden aspecten van tekstbegrip, die slechts in hun onderlinge samenhang kunnen worden geïnterpreteerd.

Vanuit de veronderstelling dat er een congruentie bestaat tussen vaardigheden van lezers en de tekstuele, propositionele niveaus, zouden we met de toetsuitslag iets aan de weet kunnen komen van het leesgedrag van de leerlingen door te kijken naar het scoreprofiel. Een één-op-één-relatie tussen vaardigheid en het niveau van de propositie waarop de stelling betrekking heeft, kan hierbij echter niet worden aangenomen. Onderzoek op het gebied van het leesproces en met name de theorie van Van Dijk en Kintsch wijzen op de voortdurende interactie tussen de verschillende verwerkingsniveaus. Het is bijvoorbeeld duidelijk dat een 'micro-propositie' uit de tekst door een 'bottom-up'-leesstrategie van de leerling kan worden geconstrueerd, maar uit het onderzoek op het gebied van lezen weten we ook dat de verwerking van dit niveau ook door een 'top down'-leesstrategie tot stand kan komen.

2.2 Resultaten

2.2.1. Verschillen in tekstbegrip(deel)vaardigheden

Voor het beantwoorden van vraagstelling 1 zijn de resultaten van de leerlingen op de verbale en nonverbale toetsen alsmede op de tekstbegriptoets geanalyseerd. Wat betreft de tekstbegriptoets gaat het hier om totaalscores (72 items); de analyse in deelscores volgt in 2.2.2. De gemiddelde totaalscores en standaarddeviaties per groep worden gepresenteerd in Tabel 1.

We zien dat de Turkse leerlingen op de talige variabelen slechter presteren dan hun Nederlandse medeleerlingen die later in hetzelfde schooltype terecht kwamen. In het algemeen presteren (toekomstige) lbo-leerlingen lager dan (toekomstige) mavo-leerlingen, zoals we kunnen

Tabel 1. Toetsresultaten (gemiddelden en standaard deviaties) van Turkse en Nederlandse mavo- en lbo-leerlingen in hun brugklasjaar.

	Turks		Nederlands	
	Lbo (n=54)	Mavo (n=34)	Lbo (n=57)	Mavo (n=32)
	Gem./SD	Gem./SD	Gem./SD	Gem./SD
Linguïstische variabelen:				
Woordkennis (max =50)	23.3/6.8	29.0/6.6	31.1/6.8	38.6/5.9
Grammatica (max= 29)	21.3/4.1	23.2/3.4	24.1/3.6	25.9/1.8
Non-verbale variabelen:				
Voorkennis (max=29)	16.1/4.9	20.7/4.5	17.0/3.0	22.0/3.1
Non Verbaal IQ(max=59)	37.0/6.8	40.3/2.4	37.7/7.4	43.1/5.1
Tekstbegrip (max=72)	47.1/6.8	52.5/6.0	49.1/6.7	56.5/5.3

verwachten. De Turkse lbo-leerlingen scoren dus van alle groepen het laagst, behalve met betrekking tot de non-verbale variabelen en tekstbegrip. Ook de Turkse mavoleerlingen doen het slechter op de linguïstische variabelen, in tegensteling tot hun relatief goede scores voor de non-verbale variabelen. Hieruit kunnen we afleiden dat er bij de Turkse leerlingen sprake is van een specifiek taalprobleem. Echter, deze taaldeficiëntie zien we niet terug in hun scores voor tekstbegrip. Het is opvallend dat de Turkse mavoleerlingen niet alleen beter dan de Nederlandse lbo-leerlingen presteren als het gaat om nonverbale toetsen, maar ook bij de tekstbegriptoets. Uit Tabel 1 kunnen we bovendien afleiden dat de verschillen in tekstbegrip tussen de twee schooltypen (lbo en mavo) groter zijn dan de verschillen tussen de twee leerlingengroepen (T1 en T2), hetgeen in Tabel 2 blijkt uit toetsing.

De resultaten van variantie-analyses (Tabel 2) laten zien dat Turkse leerlingen significant lager scoren dan de Nederlanders op alle variabelen, behalve voorkennis en non-verbaal IQ. Er is sprake van een sterk T1/T2-groepseffect met betrekking tot de linguïstische variabelen, met name woordkennis. We kunnen ook in Tabel 2 lezen dat alle effecten van Schooltype significant zijn, en zelfs sterker dan de effecten van T1- of T2 bij de non-verbale toetsen en tekstbegrip. Er is slechts een klein T1/T2-effect bij tekstbegrip. We zagen al in Tabel 1 dat de Turkse mavoleerlingen het met name relatief goed deden bij de tekstbegriptoets. Op geen van de variabelen kan evenwel een interactie-effecten waargenomen worden tussen Schooltype en T1/T2-groep.

Hoe kunnen we verklaren dat de Turkse leerlingen relatief goede resultaten behalen bij tekstbegrip ondanks hun lage scores op taalvaardigheidsmaten, met name bij de woordkennistoets? Een mogelijke verklaring is dat de Turkse leerlingen gebruik maken van top-down leesstrategieën ter compensatie van hun gebrek aan woordkennis. Ze zouden dan de neiging hebben hun problemen met de verwerking van het woord- en zinsniveau van teksten te compenseren door een relatief sterke nadruk op conceptuele en tekstuele aspecten van tekstverwerking.

Tabel 2. Variantie-analyses (2x2 ANOVA) van de toetsresultaten van Nederlandse en Turkse lbo- en mavoleerlingen in de brugklas (n.s: niet significant, p>.05).

Variabele	Effect van	M. S	DF	F	Sig. of F
Woordkennis					
	1 T1/T2	2005.3	1	43.9	.000
	2. Schooltype	1366.0	1	29.9	.000
	Interactie	5.9	1	.1	n.s.
Grammatica					
	1 T1/T2	226.0	1	18.7	.000
	2. Schooltype	88.6	1	7.3	.008
	Interactie	.1	1	.01	n.s
Voorkennis					
	1 T1/T2	28.5	1	1.6	n.s.
	2 Schooltype	628.5	1	35.4	.000
	Interactie	1.1	1	.1	n.s.
Non-verbaal IQ					
	1 T1/T2	59.4	1	1.5	n.s
	2 Schooltype	513.8	1	12.7	.001
	Interactie	33.7	1	.8	n.s
Tekstbegrip					
	1 T1/T2	214.3	1	4.8	.03
	2 Schooltype	1108.3	1	25.1	.000
	Interactie	28.0	1	.6	n.s

2.2.2 Een analyse van leesgedrag
De veronderstelling is dat er verschillen tussen leerlingen zullen optreden in hun leesgedrag, namelijk in de mate waarin ze gebruik weten te maken van 'compenserende' strategieën. Dit kan blijken uit een ander 'scoreprofiel' van bepaalde leerlingen bij de analyse van de toetsresultaten (figuur 1).

Uit deze profielvergelijking tussen anderstalige en Nederlandse leerlingen blijkt dat de eerste groep zich met name profileert op macrotekstbegrip, de Nederlandse groep daarentegen op microtekstbegrip. Beide groepen scoren slecht op de meso-subtoets. Het is voornamelijk aan hun lage microscore te wijten dat de Turkse leerlingen het slechter doen op de tekstbegriptoets. Op de andere subtests scoren ze niet significant lager dan de Nederlandse leerlingen. Het verschil tussen beide groepsprofielen betreft met name dat tussen de micro- en de macroscore: bij de anderstaligen is de gemiddelde microscore lager en bij de Nederlandse leerlingen is deze hoger dan de gemiddelde macroscore. De anderstaligen halen een bijna even grote gemiddelde macro-score als de Nederlanders, bij een gemiddeld geringere microverwerking. Hieruit zouden we kunnen afleiden dat de anderstaligen een ander leesgedrag vertonen dan de Nederlandse

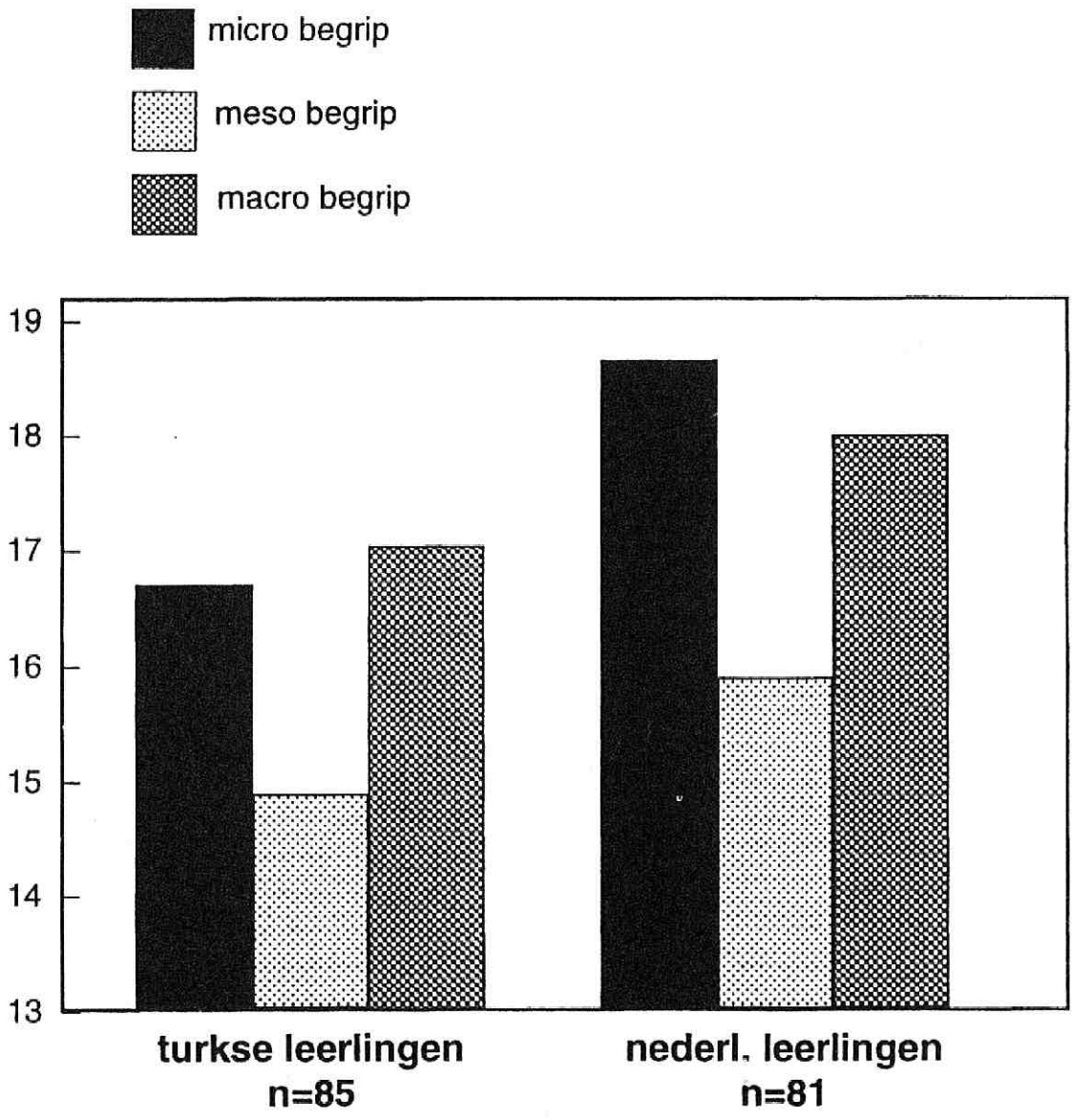


Fig. 1. Scoreprofielen van Turkse resp. Nederlandse leerlingen.

Tabel 3. Multivariate analyse (MANOVA herhaalde metingen) bij tekstbegrip op drie tekstniveaus (micro, meso and macro).

Effecten van	M.S.	DF	F	Sig. of F
1.T1/T2	234.6	1	15.7	.000
2.micro/meso/macro	243.7	2	55.8	.000
Interactie	19.9	2	4.7	.011

leerlingen. Zoals verwacht leidt hun taalprobleem, met name hun ontoereikende woordkennis, tot een 'top down'-leesstrategie. De hierboven gepresenteerde resultaten laten zien dat de Turkse leerlingen een relatief sterkere 'macroverwerking' vertonen, hetgeen indicatief is voor een 'top-down'-leesstrategie. Tabel 3 vermeldt de resultaten van de statistische toetsing van dit verschil in scoreprofielen.

Tabel 3 laat een significant interactie-effect tussen begripsniveau en T1/T2-Groep zien. Deze interactie beschrijft de tegengesteld gerichte relaties tussen de micro- and macro-subtoetsen bij de Turkse en de Nederlandse groep, hetgeen wijst op verschillende verwerkingsstrategieën van T1 en T2-lezers.

Deze profielanalyse versterkt de idee dat de T2-lezers in dit onderzoek een sterkere neiging hebben tot een top-downstrategie bij de verwerking van teksten dan de T1-lezers. Zij moeten wel sterker leunen op conceptuele sturing wegens hun problemen met de verwerking van de tekstuele microstructuur. In dit opzicht lijkt deze verwerkingsstrategie op die van zwakke lezers in het algemeen (Stanovich, 1980). Bij de T2-lezers is dit evenwel niet een gevolg van een leesprobleem, maar van een taalprobleem. De T2-lezers in dit onderzoek laten een leesgedrag zien dat potentieel bewuster en mogelijk ook efficiënter is dan het leesgedrag van T1-lezers. We zouden kunnen stellen dat, voorzover we het tekstbegrip van T2-leerlingen als problematisch beschouwen, het vooral een taalprobleem is, maar niet alleen een taalprobleem. Voorzover we het tekstbegrip van alle leerlingen als problematisch beschouwen, dan moeten we het vooral zien als een gebrek aan leesvaardigheid in de zin van strategisch gedrag. De volgende paragraaf gaat over een programma dat gericht is op een algemene verbetering van de leesvaardigheid van leerlingen.

3. TEKSTBEGRIP VERBETEREN: EEN ONDERWIJSEXPERIMENT

In een vervolgproject op dit tekstbegriponderzoek ging het om de verbetering van tekstbegrip in het voortgezet onderwijs, met name van zwakke lezers in het algemeen. Dit project heeft ten dele het karakter van een onderwijsexperiment: een groep 'leeszwakke' brugklasleerlingen (n=87) van zes scholen voor voortgezet onderwijs heeft een speciaal ontwikkeld onderwijsprogramma voor begrijpend lezen ("Weet wat je leest"-experimentversie) gevolgd. Het programma duurde tien weken. Wekelijks werden er twee lesuren aan besteed.

De groep deelnemende brugklasleerlingen is heterogeen. Verschillende groepen anderstaligen hebben de cursus gevolgd in de aparte steunlessen voor Nederlands als tweede taal, op andere scholen werden leerlingen anders gegroepeerd en kwamen Nederlandse én anderstalige leerlingen bijeen in remediërende steunlessen specifiek gericht op de verbetering van het lezen. Deze remediale hulp richt zich ook op dyslectische leerlingen. Op één school is het programma voor een deel uitgevoerd in de reguliere brugklas in het kader van studielessen. Voor alle experimentleerlingen geldt dat ze in aanmerking kwamen voor extra ondersteuning op het gebied van het begrijpend lezen volgens de op hun school geldende criteria voor selectie en plaatsing.

3.1 Het leesprogramma

We gaan ervanuit dat allochtone én autochtone leerlingen gebaat zijn met eenzelfde didactiek, waarbinnen gedifferentieerd kan worden al naar gelang de aard van het 'leesprobleem'. Het algemene doel van de cursus "Weet wat je leest" (experimentversie) is het vergroten van de functionele leesvaardigheid en het uitbreiden van de voor tekstbegrip relevante taalvaardigheid bij (leeszwakke) autochtone en allochtone leerlingen in de brugklas van het lbo en mavo.
De uitgangspunten van de experimentele cursus zijn:

1. het leren toepassen van leesstrategieën op verschillende tekstuele niveaus bij teksten van verschillende moeilijkheidsgraad;
2. het stimuleren van de reflectie op het eigen leesgedrag, het bevorderen van een bewust en zelfcontrolerend leesgedrag;
3. het bevorderen van het zelfvertrouwen en de leesmotivatie door te vertrekken bij wat de leerlingen al kunnen;
4. het aanbieden van authentiek leesmateriaal, representatief voor wat leerlingen in hun eigen omgeving aan teksten tegenkomen;
5. de integratie van taal- en leesonderwijs: het taalleren wordt bevorderd door het lezen.

De opbouw van de cursus is van 'top-down' naar 'bottom-up' en weer terug. In de eerste lessen wordt het zogenaamde 'macrolezen' geoefend. De leerlingen leren dat ze al heel wat van een tekst kunnen begrijpen zonder de inhoud nauwgezet te lezen. 'Top-down'-leesstrategieën worden in het begin sterk benadrukt, zoals bijvoorbeeld het leren benutten van de context om woordbetekenissen te raden, het inschakelen van de voorkennis en het benutten van kennis van tekststructuren: wat voor informatie geeft de 'buitenkant' van de tekst (de titel, tussentitels, alinea-indeling, opmaak e.d.). De leerlingen worden in deze eerste lessen gestimuleerd om risico's te durven lopen bij bijvoorbeeld het raden van woorden, maar ook om stukken in de tekst over te slaan wanneer die op het eerste gezicht niet van belang lijken. Het belang dat een leerling heeft bij een bepaalde tekst wordt benadrukt evenals de autonomie van de lezer die zelf beslist hoe een tekst wordt aangepakt, afhankelijk van het leesdoel.

In het vervolg van het programma wordt er 'in de tekst afgedaald': die stukken van een tekst die van belang lijken worden nader bekeken op alineaniveau. Op dit alineaniveau van een tekst speelt veel wat voor leerlingen bijzonder lastig is: het gaat hier niet zozeer om de inhoudswoorden op het microniveau maar om de functiewoorden als verbindingswoorden, pronomina en presupposities. Het gaat hier om verbanden die in de tekst gelegd worden door middel van dergelijke linguïstische elementen. We noemen dit begripsniveau 'meso'. In de methode wordt deze mesoverwerking zoveel mogelijk benaderd vanuit de tekst in zijn geheel, ook om te voorkomen dat leerlingen verzanden in een vorm van 'close reading' van een stuk tekst dat ze beter hadden kunnen overslaan. Telkens weer moet de leerling zich de vraag stellen: heb ik dit nodig, moet ik dit weten? Deze zelfbewuste en zelfcontrolerende leeshouding moet langzaam worden ingeslepen.

Het programma is uitgevoerd door docenten van de scholen zelf, die hiervoor speciale taakuren hadden. De docenten zijn door de onderzoekers en auteurs van het programma in een aantal bijeenkomsten 'geschoold' in de didactiek en vertrouwd gemaakt met het lesmateriaal, dat in de experimentversie bestond uit twintig nauwkeurig uitgeschreven lesdraaiboeken met bijbehorend teksten en oefenmateriaal.

3.2 Effecten van het programma

3.2.1 Afhankelijke variabelen

De experimentleerlingen zijn van te voren getoetst op tekstbegrip. Deze toets is dezelfde als die gebruikt is in het hiervoor vermelde onderzoek. Daarnaast is een specifiek instrumentarium ontworpen voor voor- en nameting, namelijk van die tekstbegrip(deel)vaardigheden die speciale aandacht krijgen in het leesprogramma: de zogenaamde 'Mesotoets' en 'Macrotoets'. De resultaten van de twee groepen leerlingen op deze variabelen in de voor- en nameting staan in Tabel 4.

Tabel 4. Gemiddelden en Standaarddeviaties van Nederlandse en anderstalige leerlingen op toetsen voor Tekstbegrip, Meso- en Macrovaardigheid bij voor- en nameting. T-waarden en significanties betreffen paarsgewijze t-tests per groep en per variabele.

	voortoets		natoets		T-value	Sign
	M	SD	M	SD		
T1 Tekstbegrip	50.7	5.9	52.4	5.0	1.2	.03
T2	46.5	7.0	50.2	7.0	3.8	.001
T1 Mesotoets	22.6	4.5	25.1	4.8	4.5	.000
T2	18.7	5.2	22.9	5.1	4.3	.000
T1 Macrotoets	3.1	1.4	3.9	1.4	3.7	.000
T2	3.0	1.3	3.7	1.7	1.7	n.s.

3.2.2. Tekstbegrip

Hieronder zal eerst worden ingaan op de resultaten van verschillende groepen leerlingen op de algemene tekstbegriptoets, die voor en na de cursus is afgenomen. De voor- en natoets betrof dezelfde tekstbegriptoets die hierboven reeds is beschreven. In figuur 2 zien we de vooruitgang-

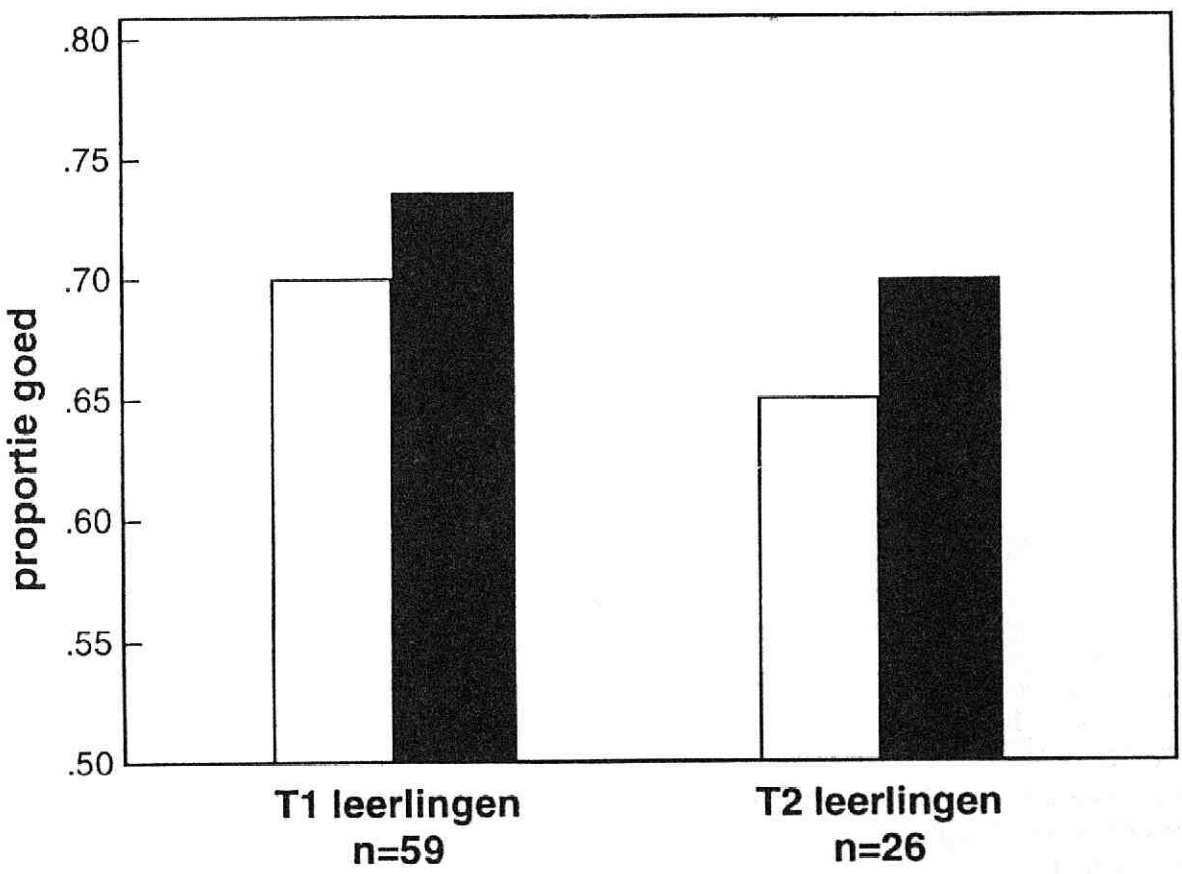


Fig. 2. tekstbegripscores van Nederlandse en anderstalige leerlingen voor en na de cursus.

scurves van de 87 experimentleerlingen die het programma hebben gevolgd en aan de voor- en nameting hebben meegedaan (59 Nederlandse en 28 anderstalige leerlingen).

We zien dat beide groepen leerlingen, zowel de anderstaligen als de (zwak lezende) Nederlandse leerlingen na de cursus gemiddeld vooruitgegaan zijn. De vooruitgang is bij de Nederlanders niet spectaculair, maar wel significant (a = .03); de allochtonen gaan duidelijker vooruit (.001). Het na de cursus bereikte leesniveau is bovendien bij alle groepen op of voorbij de door ons gestelde norm van . 70, het gemiddelde niveau van de referentiegroepen uit het eerdere onderzoek op het eind van de lbo/mavobrugklas (vgl. Tabel 1). Deze norm is tamelijk arbitrair en aan de lage kant gesteld aangezien we deze referentiegroepen evenmin kunnen bestempelen als 'goede' lezers. De vooruitgang is echter evident wanneer we bedenken dat er bij deze groepen over jaren heen geen vooruitgang wordt geboekt in tekstbegripscores bij gebruik van hetzelfde toetsmateriaal bij dezelfde typen leerlingen (Hacquebord, 1989). Wanneer de experimentgroepen door interventie van een leesprogramma wel vooruitgaan in tweeëneenhalve maand, is dat zeker een positief resultaat.

3.2.3. Meso- en macrovaardigheid

Zowel aan mesovaardigheden als aan macrovaardigheden is in de cursus expliciete aandacht geschonken. Een vooruitgang van leerlingen op deze vaardigheden moeten we dan ook als een effect van de cursus beschouwen. Ze zijn getoetst met 'leerstofspecifieke' toetsen: de Mesotoets, waarin met name op talige tekstverschijnselen (cohesie, referentie e.d.) een beroep wordt gedaan en de Macrotoets, waar het gaat om 'overall' tekstbegrip (zoals het kunnen vatten van de hoofdgedachte, het kunnen onderstrepen van kernwoorden per alinea). Beide toetsen zijn ten behoeve van het project ontwikkeld en beschreven in Galema (1989) en niet te verwarren met de in 2.2. genoemde deeltoetsen van de tekstbegriptoets. Ze zijn als afzonderlijke toetsen geconstrueerd, omdat we de beheersing van de deelvaardigheden als zodanig wilden vaststellen, aan de hand van nieuw en afzonderlijk tekstmateriaal. Op deze manier kon een grotere toetsbetrouwbaarheid worden bereikt en is interferentie tussen de deeltoetsen geminimaliseerd, waardoor de eerder (in 2.1.1) genoemde bezwaren worden weggenomen. Als we kijken naar de vooruitgang van de verschillende groepen op deze leerstofspecifieke toetsen Meso en Macro dan blijkt dat beide de groepen vooruitgaan, zij het dat de effecten verschillend in omvang zijn (Tabel 4).

Op de Mesotoets gaan beide groepen leerlingen significant vooruit (figuur 3). Bij beide groepen is sprake van een duidelijke vooruitgang, met als gevolg evenwel dat de anderstaligen hun aanvankelijke achterstand t.o.v. de Nederlandse leerlingen blijven behouden. De significante vooruitgang van de anderstaligen geldt voor alle subgroepen die te onderscheiden zijn binnen de groep anderstaligen (Turken, Marokkanen en Surinamers).

Op de Macrotoets (figuur 4) gaan alleen de Nederlandse leerlingen significant vooruit. Opvallend is dat de allochtone leerlingen bij voor- en nameting ongeveer even hoog scoren als de Nederlandse leerlingen. Bij een uitsplitsing naar verschillende etnische groepen blijkt dat met name de Turkse leerlingen (n=11) niet vooruitgaan; ze scoorden op de Macrotoets bij de pretest reeds hoog (hoger zelfs dan de Nederlanders). De Marokkanen (n=11) en met name de Surinamers (n=6) gingen daarentegen wel significant vooruit op de Macrotoets.

Deze analyses geven opnieuw steun aan de conclusie dat T2-leerlingen relatief goed presteren op macro-leestaken. Hun tekstbegripproblemen liggen met name bij de verwerking van de talige tekstelementen, hetgeen met name bij Meso-leestaken aan de orde is. Een verbetering van hun tekstbegrip op dit niveau wordt weliswaar bereikt, maar daarmee lopen zij hun achterstand ten opzichte van de Nederlanders niet in. De Nederlanders blijken met name te profiteren van de specifieke inzet van de leescursus, namelijk een verbetering op zowel de Meso- als de Macrotoets. Als het gaat om de algemene verbetering van het tekstbegrip zoals getoetst met de integrale tekstbegriptoets, valt met name de vooruitgang van de anderstalige leerlingen op. Dit is zeker

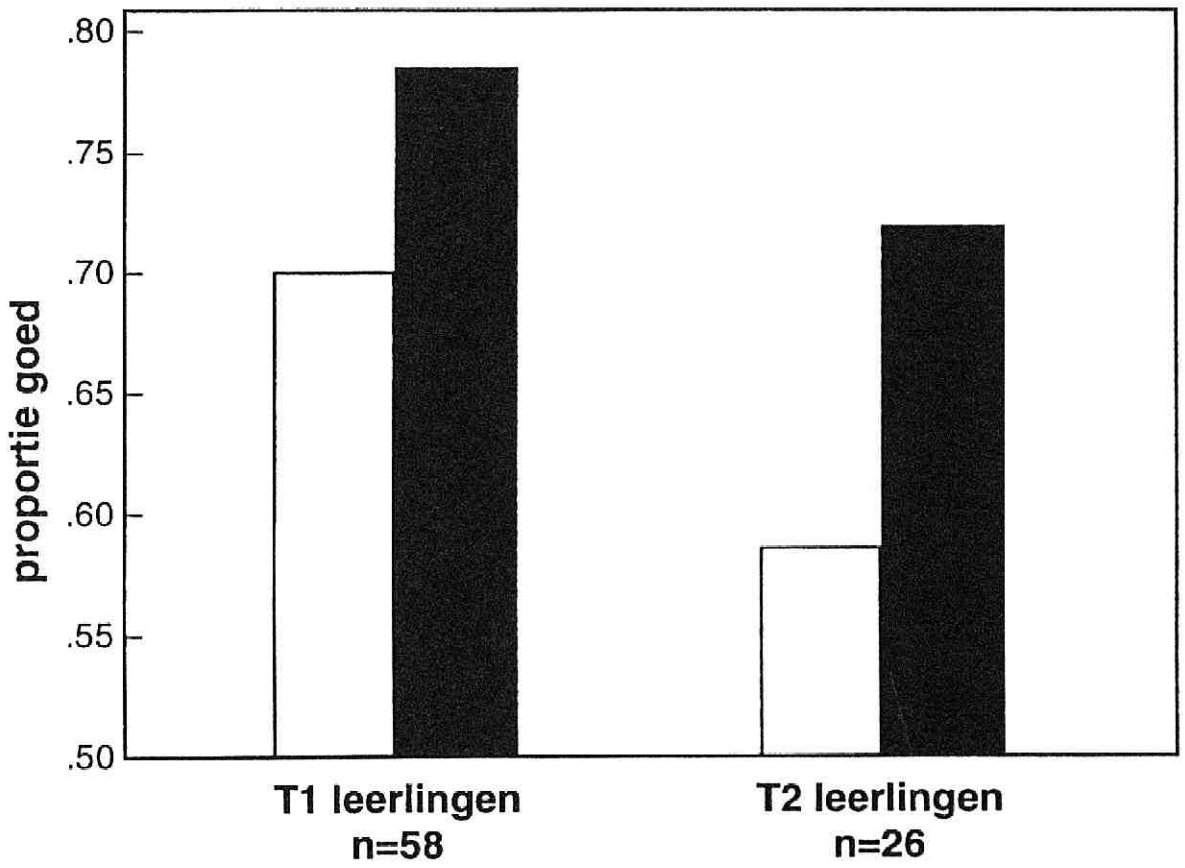


Fig. 3. Voor- en nameting Mesotoets (max.= 32) bij Nederlandse en anderstalige leerlingen.

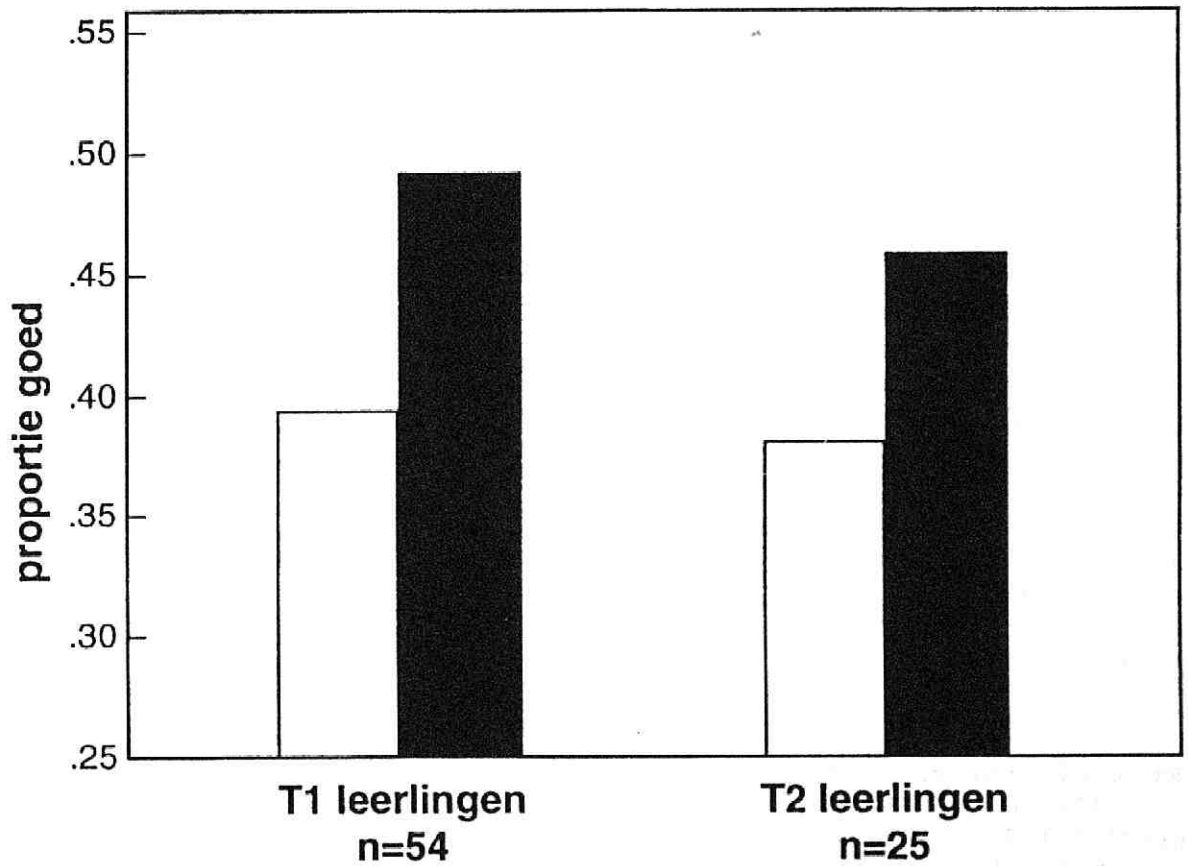


Fig. 4. Voor- en nameting Macrotoets (max.= 8) bij Nederlandse en anderstalige leerlingen . n.s.: niet significant.

een bemoedigend resultaat dat als effect van de leescursus in z'n geheel kan worden gezien. Een specifieke verbetering voor de anderstaligen, namelijk op het punt van hun voor tekstbegrip relevante taalvaardigheid, kon evenwel niet worden vastgesteld.

4. CONCLUSIES EN DISCUSSIE

Hierboven werden de resultaten van twee onderzoeken gepresenteerd, die beide tot conclusies kunnen leiden met betrekking tot het gebruik van leesstrategieën door verschillende groepen leerlingen in het voortgezet onderwijs die zich gesteld zien voor de taak teksten uit schoolboeken te verwerken. Het eerste onderzoek naar tekstbegrip bij Turkse en Nederlandse leerlingen leidt tot de conclusie dat de Turkse leerlingen sterker dan de Nederlandse leerlingen leunen op hun nonverbale vermogens als het gaat om tekstbegrip. Ze compenseren hun gebrek aan woordkennis door top-down leesstrategieën, bijvoorbeeld door een sterkere gerichtheid op de macrostructurele betekenis van de tekst. Uit ander onderzoek blijkt dat ook dyslectische lezers een dergelijk top-down leesgedrag vertonen, bij deze leerlingen als gevolg van een tekortschietende decodeervaardigheid (Hacquebord, 1991). Deze onderzoeksresultaten zijn in overeenstemming met wat o.a. Stanovich (1980) heeft opgemerkt m.b.t. het leesgedrag van zwakke lezers in het algemeen: hun tekortschieten op 'de bodem' van de tekst – in dit geval een belemmerde woordherkenning – compenseren zij (deels) door het begrip van hogere tekstuele niveaus.

De conclusie dat zwakke lezers in het algemeen geneigd zijn tot 'top-down' leesgedrag heeft geleid tot de opzet van het tweede onderzoek, waarin zwakke lezers in het algemeen, onder wie anderstaligen en dyslectici, werden betrokken. In dit onderzoek werden de mogelijkheden geëxploreerd om tot verbetering van het tekstbegrip van deze leerlingen te komen vanuit hun verondersteldc sterke kant, het hanteren van zogenaamde 'macrostrategieën'. Over de resultaten kunnen we redelijk positief zijn: het blijkt dat alle leerlingen vooruit zijn gegaan op de algemene tekstbegriptoets. Bovendien zijn de Nederlandse leerlingen specifiek vooruitgegaan op deelaspecten waarop zij eerder tekortschoten: zij gingen vooruit op de specifieke Meso- en Macrotoetsen. De anderstaligen gingen eveneens vooruit op deze specifieke toetsen, maar blijven een achterstand vertonen ten opzichte van de Nederlandse leerlingen.

Door de veldexperimentele opzet en het verkennende karakter kunnen de conclusies van dit tweede onderzoek slechts heel voorzichtig worden gepresenteerd. Ten eerste is het onduidelijk in hoeverre de vooruitgang van de experimentleerlingen te danken is aan het interventieprogamma. De referentiegroep waarmee werd vergeleken is niet afkomstig van dezelfde scholen en werd betrokken uit de steekproef van het eerste onderzoek, zodat een goede vergelijkingsmogelijkheid ontbreekt. Bovendien blijkt de experimentgroep sterk heterogeen onder andere doordat er verschillende groepen anderstaligen in werden opgenomen. Ten slotte bleek de betrouwbaarheid van het gebruikte zogenaamde "Macrotoets" te gering om conclusies over het gebruik van leerlingen van deze specifieke leesstrategie te kunnen trekken. Ook het experimentele programma "Weet wat je Leest" zou verbeterd moeten worden. Het programma lijkt vooral succesvol doordat leerlingen algemene strategische leesvaardigheden leren benutten waardoor hun tekstbegrip in het algemeen verbetert. Een specifieke verbetering voor bijvoorbeeld anderstaligen met betrekking tot de specifieke taalvaardigheidsaspecten, zoals gemeten met de Mesotoets, kon niet worden vastgesteld.

LITERATUUR

Alderson, J.C. (1984). Reading in a foreign language: is it a language problem or a reading problem? In J.C. Alderson & A.H. Urquhart (eds.). *Reading in a foreign language.* London: Longman.

Bernhardt, E.B. (1991). A psycholinguistic perspective on second language literacy. *AILA-Review 8*, 31-45.

Bossers, B. (1993). *Reading in two languages. A study of reading comprehension in Dutch as a second language and in Turkish as a first language.* Dissertatie Vrije Universiteit Amsterdam.

Brown, A.L. (1980). Metacognitive development and reading. In R.J.Spiro, B. Bertrum, W. Brewer, *Theoretical issues in reading comprehension.* Hillsdale, NJ: Erlbaum.

Brown, A.L. (1985). Metacognition: the development of selective attention strategies for learning from texts. In H. Singer & R.B. Rudell (eds.). *Theoretical Models and Processes of Reading.*

Dijk T.A. van & W. Kintsch (1983). *Strategies of Discourse comprehension.* New York: Academic Press.

Galema, C. (1988). *Tekstbegrip getoetst. Beschrijving van de constructie van toetsen voor tekstbegrip, voorkennis, woordenschat en grammatica voor de onderbouw van het voortgezet onderwijs.* Publikaties van de Vakgroep Taalbeheersing, Rijksuniversiteit Groningen.

Galema, C., Hacquebord, H., & Lijmbach, B. (1989). *Verbeteren van tekstbegrip; een onderzoek naar de effecten van een leesvaardigheidscursus voor brugklasleerlingen die problemen hebben met tekstbegrip.* Publikaties van de Vakgroep Taalbeheersing Rijksuniversiteit Groningen.

Glopper, K. de & M.E. Otter (1993a). *Nederlandse leesprestaties in internationaal perspectief.* Amsterdam: SCO

Glopper, K. de & M.E.Otter (1993b). Internationaal leesonderzoek: een bijdrage aan de discussie. In *RAiN-bulletin 4,* (3).

Hacquebord, H. (1989). *Tekstbegrip van Turkse en Nederlandse leerlingen in het voortgezet onderwijs.* Dissertatie Rijksuniversiteit Groningen. Dordrecht: Foris.

Hacquebord, H. (1991). Diagnose en behandeling van problemen met begrijpend lezen in het voortgezet onderwijs. In K. Van den Bos, H.Nakken (red.). *Dyslexie '91* p. 115-131 Amsterdam/Lisse: Swets & Zeitlinger.

Hofmans-Okkes I. (1987). *Schoolboeken leren lezen.* Muiderberg: Coutinho.

Jetten, M. (1991). De contextinvloed van scripts op woordherkenning bij tweede-taalleerders van het Nederlands. Doctoraalscriptie Toegepaste Taalkunde, Katholieke Universiteit Nijmegen.

Jong, E. de (1990). *Schoolse Taalvaardigheden in zaakvakken; leergangen en lespraktijk.* Enschede: SLO.

Oostendorp, H. van (1988). *Regulatieprocessen bij tekstverwerking.* Dissertatie, Amsterdam

Stanovich, K.E. (1980). Towards an interactive- compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly 16,* 32-71.

Stanovich, K.E. (1986) Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly 15,* 10-29.

Ulijn, J.M. & J.B. Strother (1990). The effects of syntactic simplification on reading EST Texts as L1 and L2. *Journal of Research of Reading 13* (1), 38-54.

Manuscript ontvangen 16-5-1994
Definitieve versie ontvangen 22-11-1994

APPENDIX

Voorbeeld uit Tekstbegriptotets (Hacquebord/Galema, 1989)

China
De eerste indruk

Een enorme mensenmassa, samengepakt in dalen en delta's. Van elke vijf mensen op de wereld woont er één in China. Als je nu bedenkt dat China even groot is als Europa, betekent dat dat op een oppervlakte zo groot als Europa 1/5 deel van de wereldbevolking te vinden is (In Europa woont 18% van de wereldbevolking).

Die mensen zijn zeer ongelijk over het land verdeeld. Het westen is leeg, het oosten zoveel te voller. Dat wil niet zeggen dat in het oosten elke vierkante meter bewoond is; ook hier is een scherpe tegenstelling tussen de lege bergen en de volle dalen.

China is een land van boeren. Van de beroepsbevolking werkt 80% in de landbouw. Elk stukje grond dat maar even geschikt is om bewerkt te worden is ook inderdaad agrarisch in gebruik. Wanneeer een gebied leeg blijft is dat een bewijs dat het onbruikbaar is voor de landbouw; of te steil, of te droog of te koud.

Sinds 1949 staat China, afgezien van het eiland Formosa, onder communistisch bewind. Het is naast de Sovjet-Unie de grootste communistische macht ter wereld. Het land leidt een zeer geïsoleerd bestaan, want het leeft niet alleen met de westerse wereld op gespannen voet, maar ook met de Sovjet-Unie. Dat isolement heeft tot gevolg dat we betrekkelijk weinig over het land weten, omdat sinds 1961 geen officiële cijfers, van welke aard dan ook, meer gepubliceerd zijn. We weten zelfs niet met zekerheid hoeveel inwoners China heeft.

toets 1, tekst 4 (moeilijkheidsgraad 2) Uit: Geo Geordend, deel 2m, p. 114)

Voorbeeldvragen bij de de tekst China:

microstelling:	klopt	klopt niet
1. Een 'indruk' betekent een 'aanval, duw' (microstelling)		
2. In China wont het grootste deel van de wereldbevolking (mesostelling)		
3. China is een onvruchtbaar land (macrostelling)		

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 3, pp. 271-284

Ontwikkelingen in het omgaan met literaire complexiteit

Els Andringa*
Instituut voor de Lerarenopleiding, Universiteit van Amsterdam

> Sinds de wetenschap geen vertrouwen meer stelt in algemene verklaringen en oplossingen die niet strikt afgebakend en specialistisch zijn, bestaat de grote uitdaging voor de literatuur in de kunst om de verschillende kennisgebieden en de verschillende stelsels samen te vlechten tot een meervoudige, veelzijdige visie op de wereld.
>
> Italo Calvino (1991)

ABSTRACT

Developments in the cognitive abilities of dealing with complex literary texts have hardly ever served as a guideline for the teaching of literature. Current research in text processing, however, offers insights that might be useful for the practice of literary education. In this article think-aloud protocols of expert and non-expert readers commenting on a story of William Faulkner with several narrative levels are compared. Less experienced readers were inclined to reduce the levels and to build a representation of the story-events from one perspective only. More experienced readers were able to keep different levels in mind simultaneously and to relate different perspectives. These tendencies became visible in the ways the readers represented the story's content verbally, in the causal inferences they made, in their meta-comments, and in their ways of evaluating the story.

INLEIDING

In het moderne literatuuronderwijs hebben zich verschillende opvattingen laten gelden. Sommige waren mede ontleend aan stromingen in de literatuurwetenschap. Structuralisme en New Criticism inspireerden tot een structuuranalytische benadering waarin sterk de nadruk ligt op de tekst. Mede als reactie daarop werd de receptieesthetica aangegrepen voor de ontwikkeling van een leerlinggerichte aanpak, die de responsmogelijkheden van de lezer tot uitgangspunt neemt. Daarnaast heeft zich altijd de gedachte van overdracht van het literaire erfgoed met de literatuurgeschiedenis als leidraad gehandhaafd.[1] Steeds is gezocht naar nieuwe manieren van ordenen en presenteren van werken, auteurs, bewegingen en thema's. Veel is nagedacht over de keuze van werken, over de balans tussen noodzakelijk geachte kennis en motivering van de leerling. Telkens weer zijn vragen en opdrachten geformuleerd. Echter, maar zelden heeft de vraag op de voorgrond gestaan wat iemand zou moeten *kunnen* om in brede zin met literatuur om te gaan en ervan te profiteren. Pas de laatste jaren doet het begrip 'literaire competentie' zijn intrede in de literatuurdidaktische discussie (Beach 1990, De Moor 1990, Coenen 1992, eerdere aanzet bij Purves 1981). Langzamerhand komt onderzoek op gang naar het repertoire van strategieën en activiteiten dat kenmerkend is voor het lezen van literatuur en naar de wijze waarop het "literaire lezen" zich ontwikkelt.

Vanuit de psycholinguïstiek en de cognitieve psychologie zijn modellen van tekstverwerking ontwikkeld en experimenteel getoetst. Met oog op experimentele manipulatie en als gevolg van een voorkeur voor kwantitatieve methoden maakte men lang vooral gebruik van geconstrueerd of bewerkt tekstmateriaal. De pogingen om met literaire teksten te werken bleven incidenteel. De laatste jaren groeit echter de belangstelling voor "natuurlijke" teksten. Daarnaast heeft de 'empirische literatuurwetenschap' zich met behulp van methoden uit de psychologie verder

Adres: Herengracht 256, 1016 BV Amsterdam.

ontwikkeld. Er zijn inmiddels verscheidene publikaties verschenen die aspecten van het specifiek 'literaire lezen' beschrijven of contrastief-experimenteel de verwerking van (literaire) fictie en non-fictie onderzoeken. Schmidt (1980/1991) heeft in zijn theoretisch kader voor de empirische literatuurwetenschap sterk de nadruk gelegd op verschillende conventies die het omgaan met literatuur van dat met andere talige communicatie afgrenzen. Hij neemt aan dat geschoolde lezers van deze tijd van literatuur accepteren of zelfs verwachten dat ze op verschillende manieren gelezen en begrepen kan worden, dat ze met andere woorden 'polyvalent' is. Daarnaast legt hij er sterk de nadruk op dat lezers literaire teksten verwerken in het bewustzijn dat ze 'fictioneel', d.w.z. niet *direct* op de werkelijkheid te betrekken zijn. Beach & Brown (1987) presenteren een hele reeks conventies die de verwerking van literatuur sturen zoals bijvoorbeeld de verwachting dat ieder detail van belang is, de gerichtheid op esthetische appreciatie van de tekst, de verwachting dat een verhaal een 'pointe' heeft, dat literatuur vaak een 'diepere betekenis' heeft. Ook zij wijzen op de 'conventie' van de polyvalentie.

Experimenteel onderzoek heeft een aantal verschillen in manieren van lezen aan het licht gebracht. Meutsch & Viehoff (1985) vonden bijvoorbeeld dat lezers bij literatuur meer en uitgebreidere inferenties maken dan bij populair-wetenschappelijke leesstof. Laszlo (1990) vond dat de vorming van visuele voorstellingen bij fictie anders verloopt dan bij non-fictie. Langer (1990) constateerde verschillende manieren van omgaan met informatie bij fictie en zakelijk-informatieve teksten: bij literatuur zijn lezers meer gericht op een "horizon van mogelijkheden" in het licht van eigen ervaringen en een algemene kijk op de wereld, terwijl zij bij zakelijke teksten de informatie telkens proberen te relateren aan specifieke kennis. Zwaan (1993) vond dat lezers naar verhouding meer tijd besteden aan het lezen van een tekst als ze denken dat deze tekst literatuur is dan wanneer ze dezelfde tekst voor een krantebericht houden. Bovendien onthouden zij meer van de talig-stilistische oppervlakte van de tekst. Over het verloop van literaire evaluatieprocessen is recentelijk experimenteel onderzoek gedaan door Dixon et al. (1993). In een aantal studies is tevens aandacht besteed aan ontwikkelingen in het literaire lezen en wordt de deur naar didactische toepassingen open gezet. Grzesik (1990) heeft een poging gedaan de centrale activiteiten van het literaire lezen op grond van bestaande literatuur af te grenzen en te verbinden met voorstellen voor onderwijsmethoden. Andere voorbeelden zijn te vinden bij Thomson (1987), Willenberg (1987), SPIEL (1989), Andringa & Schram (1990), Earthman (1992), specifiek over het begrijpen van metaforen en ironie: Winner (1988).

Vanuit de literatuurbeschouwing zijn sinds de oudheid telkens opnieuw pogingen ondernomen om te definiëren wat literatuur is en wat het literaire karakter van een tekst bepaalt. In het denken over de kenmerken van de teksten en de werking die daaraan toegeschreven werd lagen soms ook vooronderstellingen over het (ideale) lezerspubliek besloten. Pas in de jaren zeventig werd in de zogeheten *receptieesthetica* "de lezer" uitdrukkelijk onderwerp van beschouwing en kregen de literaire leesprocessen expliciet aandacht (Iser 1975). Uit deze traditie kwam niet direct empirisch onderzoek voort, wel zijn er aannames over het literaire lezen ontwikkeld. De receptieesthetici gingen in navolging van de Russische en Tsjechische structuralisten vooral uit van specifiek literaire eigenschappen in taal, structuur en genre en probeerden de wijze waarop deze de (ideale) lezer zouden sturen te reconstrueren. Zij wezen met name op verschijnselen als meerduidigheid, onbepaaldheid en 'open plekken', op gelaagdheid in betekenissen, tijd en perspectieven, op vormen van paradoxie en incoherentie, en op gevallen van indirectheid zoals metaforiek, symboliek en ironie. Volgens hun opvatting worden lezers daarbij telkens geconfronteerd met afwijkingen van normen en verwachtingen die bij "normale" communicatie gelden en worden zij tot extra verwerkingsactiviteiten aangezet. Hun constructies van leesprocessen impliceren daarbij een theoretische lezer die in staat is dergelijke vormen van literaire complexiteit 'adequaat' te hanteren. Over de werkelijke verwerkingsprocessen en over de ontwikkelingen tot dit 'adequate lezen' is echter nog steeds weinig bekend. De hypothetische leesprocessen zijn slechts bij uitzondering ook empirisch geverifieerd.

Bij deze stand van zaken wil ik in deze bijdrage aanknopen. Ik wil proberen om een klein stukje specifiek literaire tekstcomplexiteit in kaart te brengen en te laten zien hoe daarbij verschillen in de verwerkingsprocessen optreden die met literaire leesvaardigheid samenhan-

gen. Ik richt mij op een geval van 'gelaagdheid' zoals dat voorkomt in William Faulkners fameuze verhaal 'A Rose for Emily'. Het onderzoeksmateriaal dat facetten van de verwerkingsprocessen zichtbaar maakt bestaat uit 'hardop-denk-protocollen'. Tekst en methode zullen eerst worden toegelicht.

1. DE TEKST: 'EEN ROOS VOOR EMILY'

De eigenlijke handeling van Faulkner's verhaal bevat de levensloop van een eenzame, maar autoritaire dame die achteraf haar geliefde blijkt te hebben vergiftigd. Hoewel ze ooit bij de drogist arsenicum had gekocht en men gedurende enige tijd stank rond haar huis had waargenomen, ontdekt men pas na haar dood, dat ze het lijk tientallen jaren in haar slaapkamer heeft bewaard. Het verhaal wordt verteld door een ik-verteller die herinneringen ophaalt en schijnbaar associatief episodes uit verschillende tijden memoreert, gepresenteerd als dat wat hijzelf en de buurtgenoten in de loop der tijden van Miss Emily hadden waargenomen. De hoofdpersoon wordt daardoor uitsluitend van buitenaf, en bovendien terugblikkend in de tijd beschreven. Drie vertellagen spelen hier dooreen: 1. de eigenlijke handeling, d.w.z. de levensgeschiedenis van Emily, 2. het verhaal van de dorpsgenoten – waartoe ook de verteller zelf in vroeger tijd behoort – die fragmenten van het gebeuren observeren en op hun manier interpreteren en becommentariëren, en 3. het vertellend herinneren van de verhaalepisodes door de verteller. Emily's levensverhaal is dus, hoewel het de eigenlijke handeling vormt, alleen indirect, namelijk via de herinneringen van de verteller, en daarbinnen weer via de observaties van de gemeenschap toegankelijk.

Lezers zien zich nu herhaaldelijk geconfronteerd met sprongen in tijd en informatie en moeten het oorspronkelijke gebeuren uit de verschillende lagen reconstrueren.[2] Zij trachten bijvoorbeeld de chronologie te herstellen, maken inferenties over de oorzaken van de handeling, motieven en intenties van de personages, en zetten soms ook de gezichtspunten uit de verschillende "lagen" tegenover elkaar (zie Andringa & Davis 1994). Om een vergelijking tussen verschillende lezers mogelijk te maken dient in het volgende een passage tot uitgangspunt waarin de gelaagdheid duidelijk een rol speelt, en waarbij in de respons van de informanten een heel scala aan verwerkingsactiviteiten zichtbaar wordt. Het gaat om het fragment waarin Emily naar de drogist gaat om arsenicum te kopen direct gevolgd door een fragment waarin opinies van de gemeenschap naar voren komen. De informanten kregen de passage als de pagina's 9 en 10 aangeboden.

[9] Ze hield het hoofd hoog – zelfs toen we dachten dat ze diep gezonken was. Het was net alsof ze meer dan ooit eiste dat men haar waardigheid als de laatste der Griersons zou erkennen; alsof alleen dat vleugje aardsheid nodig was geweest om te bevestigen hoe onaantastbaar ze was. Zoals bijvoorbeeld toen ze dat rattengif kocht, het arsenicum. Dat was meer dan een jaar nadat ze 'Arme Emily' waren gaan zeggen, en toen de twee nichtjes haar nog kwamen opzoeken.
'Ik wil wat gif,' zei ze tegen de drogist. Ze was toen in de dertig, nog steeds een tengere vrouw, magerder zelfs dan gewoonlijk, met koude, hooghartige donkere ogen in een gezicht waarvan het vlees strak over de slapen en oogkassen was getrokken zoals je je het gezicht van een vuurtorenwachter voorstelt. 'Ik wil wat vergif,' zei ze.
'Ja, Miss Emily. Wat voor soort? Tegen ratten en zo? Dan raad ik...'
'Ik wil het beste dat u hebt. 't Kan mij niet schelen wat voor soort.'
De drogist noemde enkele op. 'Die doden alles, tot een olifant toe. Maar wat u nodig hebt...'
'Arsenicum,' zei Miss Emily. 'Is dat goed?'
'Of arsenicum...? Ja, mevrouw, maar wat u nodig hebt...'
'Ik wil arsenicum.'
De drogist keek op haar neer. Ze keek terug, kaarsrecht, het gezicht als een strak gespannen vlag.
'Maar natuurlijk,' zei de drogist. 'Als u dat wilt. Maar volgens de wet moet u zeggen waarvoor u het gaat gebruiken.'
Miss Emily keek hem alleen maar strak aan, het hoofd schuin achterover om hem oog voor oog aan te kijken tot hij zich van haar afwendde en het arsenicum ging halen en het inpakte. Het zwarte loopjon-

getje bracht haar het pakje; de drogist liet zich niet meer zien. Toen ze het pakje thuis openmaakte stond op het doosje onder de schedel met knoken: 'Voor ratten.'

[10] Dus zeiden we de volgende dag allemaal: 'Ze gaat zelfmoord plegen'; en we besloten dat dat maar het beste was. Toen we haar voor het eerst met Homer Barron zagen, hadden we gezegd: 'Daar gaat ze mee trouwen.' Toen zeiden we tegen elkaar: 'Ze krijgt hem nog wel zover,' omdat Homer zelf had opgemerkt – hij was graag onder mannen en het was bekend dat hij met de jongelui in de Elks' Club dronk – dat hij er de man niet naar was om te trouwen. Later zeiden we, van achter de jaloezieën: 'Arme Emily' als ze zondagmiddag in het glimmende rijtuig langskwamen, Miss Emily met opgeheven hoofd en Homer Barron, de rand van zijn hoed opgeklapt en een sigaar tussen zijn tanden, teugels en zweep in een gele handschoen.
Toen begonnen sommige dames te zeggen dat het een schande voor het stadje was en een slecht voorbeeld voor de jeugd. De mannen wensten zich er niet mee te bemoeien, maar tenslotte dwongen de dames de dominee van de baptistengemeenschap – de familie van Emily was Anglicaans – haar een bezoek te brengen. Hij heeft nooit willen vertellen wat er tijdens dat onderhoud was voorgevallen, maar hij weigerde er nog eens heen te gaan. De volgende zondag reden ze weer door de straten en de dag daarop schreef de domineesvrouw naar de familie van Miss Emily in Alabama.

De passage is karakteristiek voor de manier van vertellen: een stukje uit Emily's geschiedenis wordt gevolgd door een stukje commentaar van diverse partijen uit de gemeenschap dat wederom brokstukjes informatie over Emily bevat. Binnen deze passage treden verschillende sprongen op in de (herinnerde) tijd en de informatie, gekleurd door de visie van de omwonenden. Te verwachten is dat hier een spanning ontstaat of versterkt wordt tussen het beeld van het gebeuren ingebed in de voorstelling van de gemeenschap en het beeld dat de lezer zelf opbouwt van Emily en van wat er "echt" is gebeurd. De vraag is nu wat lezers doen met de complicaties die de manier van vertellen met zich meebrengt en hoe lezers met meer en minder leeservaring daarin verschillen.

2. METHODE

Het verhaal werd ingedeeld in 13 passages die inhoudelijk gezien een relatief afgeronde eenheid vormden: ze bestonden uit een of meer onderscheidbare handelingsepisodes. De passages werden de informanten stuk voor stuk in de originele volgorde voorgelegd. De instructie luidde om telkens na lezing van ieder stuk zo direct mogelijk "hardop te denken" en daarbij alles te zeggen wat naar aanleiding van het stukje was opgekomen respectievelijk alsnog opkwam. Deze methode registreert weliswaar niet de on-line processen, maar vangt wel de secundaire verwerking op: de reflectie naar aanleiding van het gelezene, de registratie van begripsproblemen, de terugkoppelingsprocessen naar wat eerder gelezen is, en de vorming van associaties en verwachtingspatronen. Daarnaast gaven veel lezers spontaan affectieve reacties en eerste oordelen over zowel personages en handeling als over tekst of stijl.[3] Ook reflectie op het eigen leesproces trad regelmatig op. Na de hardop-denk-procedure voerden proefleider en informant een gesprek over het verhaal. Rode draad door het gesprek vormde een aantal vaste vragen. Verder was de vorm van het gesprek zo vrij mogelijk. De eerste vraag was in de meeste gevallen wat men van het verhaal als geheel vond. Naast de reacties op het boven geciteerde fragment uit het verhaal zullen we ook aan de respons op deze vraag aandacht schenken. In het gesprek werd tevens informatie vergaard over leesgewoonten en leesinteresses van de informanten. Alle protocollen en gesprekken werden opgenomen en getranscribeerd. Het verhaal werd zowel in de originele versie aan Amerikaanse, als in de Nederlandse vertaling aan Nederlandse respondenten voorgelegd. De Engelstalige reacties zijn hier in het Nederlands weergegeven.

Het onderzoek had uitdrukkelijk een exploratief karakter. Het was niet de bedoeling om hypothesen over de invloed van lezerskenmerken op leesprocessen te toetsen, maar om deze te verkennen. Daarom werden relatief verschillende lezers benaderd. Er zijn echter wel groeperingen van de informanten mogelijk. In totaal namen 22 informanten van verschillende leeftijd en met een verschillende achtergrond qua opleiding deel. Van hen waren 14 Amerikaan, 8 Neder-

lander. Eventuele cultuurverschillen laten we hier buiten beschouwing. 12 Informanten waren vrouw, 10 man; ook op verschillen in sexe gaan we hier niet in. We zullen echter wel aandacht schenken aan de achtergrond van de informanten voorzover deze met leeservaring en -interesse te maken heeft. We onderscheiden hier drie groepen, namelijk: 7 ca. 16/17-jarige scholieren ofwel highschool students zonder speciale leeservaring en -interesse (SCHOL), 8 niet-letterenstudenten en andere jonge volwassenen zonder speciale leeservaring en -interesse, leeftijd tussen 20 en 30 jaar (VOLW), en tenslotte 7 ouderejaars letterenstudenten en jonge volwassenen met relatief veel leeservaring (ERVAR), ook hier lag de leeftijd tussen 20 en 30 jaar. In dit artikel worden bij ieder aangehaald geval een voornaam en de groepscode vermeld. Bij de hier gepresenteerde analyse van het materiaal spelen dus een rol: a. verschillen in leeftijd en de daarmee verbonden "rijpheid" en "levenservaring" (die ongetwijfeld een rol spelen bij het lezen van literatuur) en b. verschillen in leeservaring.

De protocolfragmenten van informanten worden vergelijkend geanalyseerd. Daarbij wordt geprobeerd om verschillen in de verwerkingsstrategieën en -activiteiten tussen 16/17-jarige "doorsnee" scholieren, "doorsnee" jonge volwassenen en "expert-lezers" aan te wijzen. Gekeken wordt in eerste instantie op welke "laag" of "lagen" de lezers zich in hun reacties richten en vanuit welk perspectief of welke perspectieven ze voorstellingen van de personages en handelingen presenteren. Als resultaat is een overzicht van strategieën en activiteiten beoogd dat het mogelijk maakt verschillende niveaus in het omgaan met de gelaagdheid te onderscheiden.

3. RESULTATEN EN ANALYSE

3.1. Verwerking van de verhaalstructuur. Een vergelijkende analyse

Check is een highschool leerling die weinig in lezen geïnteresseerd is, maar heel welwillend aan het onderzoek meewerkt. Zijn reactie op de onder 2 geciteerde tekstpassages 9 en 10 luidde als volgt:

Check [SCHOL]
9: Ik denk dat ze een beetje getikt is dat ze daar gif gaat kopen en ze wil niet tegen de drogist zeggen waar ze 't voor nodig heeft. Die Emily wordt daar een beetje gek.
10: Kennelijk is ze er slecht aan toe en eh, misschien is het maar het beste dat ze zichzelf van kant maakt.

Check reageert primair op de gebeurtenis die in de tekstpassage verhaald wordt. Zonder er zelf erg in te hebben laat hij zich leiden door de manier waarop de gemeenschap Emily observeert. Hij neemt hun visie op "Arme Emily" over: hij accepteert de veronderstelling dat ze zelfmoord gaat plegen en herhaalt dat "dat maar het beste is". De gelaagde constructie van het verhaal gaat hierbij verloren. De lagen worden als het ware samengeklapt tot een enkelvoudige voorstelling vanuit het gezichtspunt van de buurt.

David is wat gedetailleerder in zijn reactie:

David [SCHOL]
9: Dit stukje gaat over dat Miss E. uitgaat en arsenicum koopt en er is eigenlijk niks waar ze het voor koopt... normaal willen de drogist en de wet dat je het ergens voor koopt, maar omdat Miss E. een dame is laat de drogist het maar zitten.
10: De mensen geloven dat ze dat arsenicum had gekocht om zelfmoord te plegen. Ook dachten ze dat die noorderling, Homer Barron, haar zou kunnen redden en met haar trouwen. Maar ze twijfelden en de vrouw van de dominee schreef aan haar familie in een ander deel van de staat om te zien of ze konden komen helpen.

Deze reactie resumeert globaal de inhoud van de passage – een strategie die vooral bij scholieren voorkomt – en richt zich daarbij op het niveau van de handeling. De gelaagdheid gaat ook hier inzoverre verloren als David het handelen van Emily en het observeren en reageren van de

mensen op dezelfde lijn stelt: "Dit stukje gaat over.." leidt een homogeniserende weergave van gebeuren, observeren en vertellen in. Terwijl hij het verhaal tot één verhaalniveau reduceert, maakt David wel enkele inferenties: de drogist laat het maar zitten "omdat ze een dame is", en "ze dachten dat [...] Homer Barron haar zou kunnen redden". Ook hier blijkt de neiging om met het perspectief van de buurt mee te denken.

De respons van Kevin is minimaal, maar toch geeft hij aan dat hij zich van verschillende visies bewust is:

Kevin [SCHOL]
9: Het lijkt wel of Miss Emily een stomvervelend leven heeft. Het lijkt wel of ze verlamd is.
10: Ik geloof niet dat Miss Emily goed is, ze hadden haar niet op zo'n voetstuk moeten zetten.

De eerste reactie gaat geheel aan de gelezen passage voorbij, maar geeft een globale impressie van de hoofdpersoon weer. Ook de tweede reactie is van algemene aard, maar hier blijkt wel dat deze lezer zich een beeld van de hoofdpersoon vormt dat anders is dan het beeld dat hij aan de gemeenschap toeschrijft. Hoewel niet gerelateerd aan de gelezen passages blijkt Kevin zich een dubbele voorstelling van Emily te vormen. Zijn eigen opinie is duidelijk te onderscheiden van die van Emily's buurtgenoten.

Groepsgenote Shannon reageert uitvoeriger:

Shannon [SCHOL]
9: Ik geloof dat ze gaan verliezen of, ehm, ze gaat beslist iets doen met dat rattegif en niemand kan haar tegenhouden want ze is een heel machtige vrouw.
10: Grappig hoe die mensen verkeerd vonden wat Miss Emily deed, ik geloof dat ze eigenlijk helemaal niet veel deed, maar ze proberen een eind aan die relatie te maken. Grappig, want het is alsof ze doet wat zij niet willen dat ze doet.

Deze reactie verschilt in drie opzichten van de eerste voorbeelden. *Ten eerste* onderkent Shannon duidelijk een tegenstelling tussen haar eigen beeld van Emily's doen en laten en de opvattingen van de gemeenschap. Zij karakteriseert de relatie tussen Emily en de gemeenschap in termen van macht enerzijds, en van weerstand tegen de gangbare opvattingen anderzijds. Iets van de inbedding van Emily's geschiedenis in de visie van de gemeenschap blijft bewaard in de manier waarop zij haar zinnen opbouwt: "Grappig hoe die mensen verkeerd vonden wat Miss Emily deed" en "Grappig, want het is alsof ze doet wat zij niet willen dat ze doet". In de *tweede* plaats ligt in deze karakterisering van de verhoudingen ook een generalisatie besloten die uitgaande van de gegeven passage de positie van Emily in de gemeenschap in ruimere zin omvat ("ze is een machtige vrouw" en "het is alsof ze doet wat zij niet willen dat ze doet.") En in de *derde* plaats blijkt uit Shannon's reactie dat ze geamuseerd is door de constellatie ("Grappig"). Daarin onderscheidt ze zich duidelijk van Check en Kevin die alleen het personage van Emily evalueerden, en van David, die geheel op een reconstructie van het gebeuren gefixeerd was.

In deze selectie van reacties komen corresponderend met de vertellagen twee "beelden" van Emily naar voren: 1. Emily zoals de lezer zich een voorstelling van haar vormt los van de gemeenschap, en 2. Emily zoals ze via de verteller door de ogen van dorpsgenoten wordt gepresenteerd. Sommige lezers onderscheiden deze beelden niet, maar trekken ze samen in één voorstelling. Van de derde laag, de constructie van het geheel door de verteller, wordt in deze reacties niet direct iets zichtbaar. De oorspronkelijke vertelstructuur gaat in vorm en functie verloren.

Verder hebben we als verwerkingsactiviteiten geïnventariseerd: het maken van – meest causale – inferenties op het handelingsniveau, karakterisering van de personages, generaliserende karakterisering van situatie of personage, en het evalueren van personages en handeling. Wat *niet* in de protocollen van de scholieren voorkwam, waren verbindingen naar informatie elders uit de tekst, signalering van de sprongen in de tijd, en commentaar op de tekststructuur überhaupt. De scholieren reageren veelal van fragment naar fragment waarbij zij zich op de daarin gegeven brokstukjes handeling concentreren.

We zetten de reeks van voorbeelden voort met twee reacties uit de groep van jonge volwassenen zonder speciale leeservaring.

Jeroen [VOLW]
9: 't is een eh eh arrogante vrouw die zich boven de rest verheven voelt, die zich afzijdig houdt van iedereen en eh die zich die zich ehm meer vindt, zich er bovenuit voelt steken. 't Is allemaal erg afstandelijk voor me, ik lees het als iets heel afstandelijks. Het lijkt erop alsof ze de boel wil choqueren of dat het een manier is om aandacht te krijgen.
10: Ja de hele situatie is duidelijk om ehh de bevolking te choqueren. Er wordt geloof ik ook gesuggereerd dat die Homer helemaal niet op vrouwen valt, dat het een flikker is. Z'n z'n gele handschoentjes of staat dat er niet? Het zou een schande voor de stad zijn.

Jeroen richt zich vooral op een algemene karakterisering van Emily en probeert daarbij beweegredenen voor haar handelen te geven. Ook maakt hij de inferentie dat Homer misschien homosexueel is. Hij signaleert wel een tegenstelling van Emily en het dorp, maar verbindt de gezichtspunten op één verhaalniveau. Daarbij plaatst hij zich buiten de opinies van de gemeenschap en denkt eerder vanuit Emily: "Het lijkt alsof ze de boel wil choqueren". Een nieuw element wordt gevormd in de reflectie op zijn eigen leesproces" "ik lees het als iets heel afstandelijks". Dit is een aanwijzing dat deze lezer zich van een werking van de manier van vertellen bewust is; hij geeft echter niet aan inhoeverre hij deze werking zelf aan de vertelstructuur toeschrijft.

Elizabeth [VOLW]
9: Wat krijgen we nou? Het lijkt wel haast of het twee verschillende verhalen zijn. Het eerste ehm stuk van deze bladzij vertelt nog steeds over haar positie, maar het gaat over van een soort passieve beschrijving naar een heel actieve passage waar ze echt iets doet en eh eh ze is erg veeleisend en hooghartig, mijn sympathie voor Emily gaat steeds op en neer, ik denk haast dat ik iets van haar wil maken wat ze niet is en dat probeer ik in het verhaal leggen. En waar zou ze dat arsenicum voor nodig hebben? [lacht]
10: Dit maakt me woedend. Ik ben erg gevoelig voor mensen die roddelen, die anderen proberen te vertellen hoe ze moeten leven, om dingen die me zelf zijn overkomen. Mijn oudste zus was een ongehuwde moeder toen die term nog gebruikt werd, en ehm toen mensen daar nog echt op neerkeken en ik weet hoe wreed mensen kunnen zijn en wat ze zeggen en doen en mijn eigen moeder was geestesziek en ik weet wat mensen kletsen en doen, echt wreed. Dit doet me denken aan die aan die dames, die ... echt typisch van die enge ouwe religieuze taarten die zich overal mee bemoeien, heb ik gekend, die christelijke weldoeners, mijn tante is daar een goed voorbeeld van. Wie weet wat voor iedereen het beste is.. ik kan me daar geweldig over opwinden.

In de eerste reactie geeft deze lezeres duidelijk te kennen dat ze de twee niveaus als verschillend ervaart, ze licht het verschil toe met een karakterisering van de manier van beschrijven: passief versus actief. Daarnaast ziet zij haar eigen rol als lezer, ze onderkent haar eigen neiging tot emotionele betrokkenheid ("ik denk haast dat ik iets van haar wil maken wat ze niet is"). Desondanks raakt Elizabeth, zo blijkt uit de tweede reactie, sterk emotioneel bij het verhaal betrokken door associaties met haar eigen leven. Zij vormt zich een uitgesproken voorstelling van de buurtgenoten en oordeelt heel negatief over hun manier van denken en doen. In haar elaboratie van persoonlijke herinneringen lijkt ze de dualiteit van de tekst haast weer uit het oog te verliezen.

Hoewel Elizabeth extreem is in haar emotionele kritiek op de gemeenschap, wordt in de groep van de volwassen lezers in het algemeen meer aandacht geschonken aan de partij van de dorpsgenoten. De meesten van hen signaleren en bekritiseren de bemoeizucht en de benepen visie. Wil zegt bijvoorbeeld: "De indruk die ik nou heb is hoe benauwend het is hè om in zo'n kleine gemeenschap te wonen, waarin iedereen zo vreselijk roddelt." En Rieneke zucht: "Mensen.. wat daar allemaal niet mocht!" Deze lezers vormen zich meer dan de jongste groep een voorstelling van de gehele situatie, en onderscheiden daarin duidelijk de kijk van de gemeenschap op het gebeuren. Zij zijn daardoor over het algemeen minder op de stukjes van de eigen-

lijke handeling gefixeerd dan de scholieren. Het simpelweg overnemen van het standpunt van de buurt komt niet meer voor, ook het volledige in elkaar schuiven van de posities vinden we hier niet meer. In het generaliseren van Emily's positie, de gehele situatie en de sfeer in de gemeenschap blijft het overzicht over het verhaal als geheel beter behouden. Toch verwijst men ook in deze groep nergens precies naar informatie elders in de tekst, en ook reflectie op de wijze van vertellen of de structuur vindt hier zelden plaats. Tenslotte zij nog vermeld dat de reacties van de ouderen in het algemeen veel uitvoeriger zijn.

Tot slot twee representatieve gevallen uit de groep van ervaren lezers. Eva staat aan het eind van haar letterenstudie en heeft een uitgesproken belangstelling voor literatuur. Haar reactie was:

Eva [ERVAR]
9: Het is een beetje van de hak op de tak-achtig hè, nu beginnen ze opeens over rattengif. Dat is dan weer een heel andere tijd dat dat speelt. Dat is wel een paar keer zo gebeurd. Het is wel associatief geschreven, alsof de verteller hierdoor ergens anders op komt en dat dan weer gaat vertellen. Hm, tsja, ze legt gewoon iedereen haar wil op, hè, tot de drogist toe. Echt, eh, ik weet niet hoe het afloopt, ik weet niet wie ze gaat vermoorden, ja.
10: Hm, nou, ik had niet meteen aan zelfmoord gedacht, dat leek me niks voor haar, dat past gewoon niet bij het beeld dat ik had van haar. mmm, dus, het gaat via de dames naar de heren die dus de macht hebben en er wordt iets geregeld. Het is echt zo van "Arme Emily" zeggen dat is eh, medelijden met haar hebben dat is juist het ergste wat ze zo iemand aan kunnen doen. Want andere dingen raken haar niet.

In dit commentaar speelt inzicht in de constructie mee: Eva karakteriseert de manier van vertellen en onderkent de sprongen in de tijd. In een adem door geeft ze ook een algemene karakterisering van Emily waarbij ze de scène bij de drogist als een nieuwe illustratie van Emily's wil registreert; ze plaatst deze scène dus in een ruimer tekstverband ("iedereen ... tot de drogist toe"). Terloops maakt ze een inferentie over Emily's intentie een moord te plegen. Zoals uit de tweede reactie blijkt heeft ze een zodanig duidelijk beeld van Emily opgebouwd, dat ze de suggestie van zelfmoord direct in twijfel trekt. Vervolgens differentieert ze ook het beeld van de gemeenschap ("van de dames naar de heren die dus de macht hebben"). En tenslotte confronteert ze haar eigen beeld van Emily met dat van de gemeenschap. Noch de zelfmoord, noch het medelijden acht ze op Emily van toepassing. Voor deze lezer is de voorstelling van de gemeenschap duidelijk niet adequaat.

Mirjam's respons is nogal uitvoerig; we beperken ons tot haar reactie op het tiende fragment:

Mirjam [ERVAR]
10: Het is wel een verhaal waarin dingen pas later worden uitgelegd hè, later pas duidelijk worden.
I: Zoals?
M: Nou hoe dat met die Homer Barron is afgelopen. Het zou me verbazen als ze echt zelfmoord ging plegen, omdat het zo voorspelbaar is, maar ja.. Het is een draak van een verhaal. Ja, het is een ontzettend bedompte sfeer die eruit tegemoet walmt. Dan die dominee er nog bij. En er zit ook helemaal geen spanning in. Zoals dat niet willen onthullen, die dominee, wat er is gezegd tijdens het onderhoud. En dat ze, die vrouw schrijft naar haar familie, dat zou dan spanning moeten geven denk ik. Anders zou de schrijver nu wel zeggen wat er gezegd is, dus er wordt informatie achter gehouden. Maar eh, dat werkt bij mij niet.
I: Waarom niet?
M: Eh, ja ik weet niet hoeveel bladzijden ik nu al gelezen heb, maar er staat echt niets verrassends in. Het stuurt allemaal aan op de ondergang van die vrouw en eh, ik vind eh 't gewoon geen kunst om iemand eh, zo van buitenaf te bekijken met zo'n enorme afstand, als een karikatuur. Dat boeit me totaal niet, zo'n verhaal. Ik vind het wel boeiend om te zien hoe iemand zijn waardigheid probeert te behouden en dat lukt niet en daaraan gaat ie dan onderdoor. Maar om dat van zo'n afstand te, ja, het komt niet tot leven, laat ik het zo zeggen. Het hele personage niet. Wat ik ook vervelend vind is dat anonieme we-perspectief. Je ziet er gewoon geen gezichten bij. En daarom wordt het niet echt spannend, het is allemaal zo ver weg.

Bijna deze gehele reactie bestaat uit negatieve kritiek, en wel kritiek die enerzijds op de inhoud betrekking heeft ("bedompte sfeer"), anderzijds vooral ook op de manier van vertellen. Mirjam doorziet een bepaalde techniek, namelijk dat de tekst sommige informatie niet geeft, maar ontkent dat dit bij haar spanning oproept. Interessanter is echter haar opmerking over de verhouding van de verteller ten opzichte van zijn onderwerp: zij vindt dat wat zij als thema beschrijft ("hoe iemand zijn waardigheid probeert te behouden etc.") de moeite waard, maar bekritiseert de manier waarop dit wordt gebracht. Dat, wat Jeroen als afstandelijk omschreef, trekt zij door tot "karikatuur". Uitdrukkelijk vermeldt zij ook het "we-perspectief". Analyse van de vertelstructuur toonde aan dat het "eigenlijke gebeuren" op een dubbele manier is ingebed. Voor het eerst wordt hier in de respons iets zichtbaar van het derde niveau, dat van het vertellen door een representant van de gemeenschap. Hoewel minder getrainde lezers zich op de verhaalde handeling richten en de structuur in veel gevallen reduceren, blijkt nu bij de volwassen lezers dat de structuur wel degelijk effect heeft, zij het dat dit niet altijd positief ervaren wordt. Sommige lezers voelen een afstandelijkheid die ze het verhaal niet in dank afnemen. Het lijkt zelfs of Mirjam meer betrokkenheid van de verteller bij het personage gewenst had. Als meer ervaren lezer doorziet zij een aantal vertelkenmerken, maar kan daarvoor geen waardering opbrengen.

3.2. Samenvattend overzicht van de verwerkingsstrategieën en -activiteiten

In het materiaal dat we hier hebben getoond zijn verschillende soorten verwerkingsactiviteiten zichtbaar geworden. Sommige zijn vooral cognitieve inspanningen om het verhaal te begrijpen, andere zijn vooral evaluerend van aard en soms zelfs sterk emotioneel gekleurd. Maar bijna overal is de respons te verbinden aan de diverse lagen in het verhaal. We zullen in de volgende paragrafen nog eens beknopt weergeven waarin de duidelijkste verschillen tussen de drie lezersgroepen in het verwerken van en reageren op de gelaagdheid bestaan. Het gaat daarbij om de wijze waarop de informanten de lagen globaal waarnemen (3.2.1.), om de verschillende soorten inferenties die zij maakten (3.2.2.), en om de manieren waarop zij het verhaal evalueren (3.2.3.).

3.2.1. Waarneming van de gelaagdheid

We gaan in de eerste plaats uit van twee voorstellingen van het gebeuren, namelijk de voorstelling van Emily's persoonlijkheid en van haar levensgeschiedenis zoals de lezer die langzamerhand uit de verschillende brokjes informatie afleidt (V1) en de voorstelling van Emily's persoonlijkheid en van haar doen en laten door de ogen van de buurtgenoten (V2). Daarnaast vormen sommige lezers zich ook een voorstelling van de verteller en van hetgeen deze met zijn manier van vertellen beoogt (V3). In de gepresenteerde protocollen is een aantal manieren van verwerking de revue gepasseerd, namelijk: 1. Overname van V2 in of als V1 (Check), 2. samentrekking van V1 en V2 (David), 3. signalering van verschil V1 en V2 (Kevin), 4. op elkaar betrekken van V1 en V2 (Shannon), 5. signalering van de verhaalstructuur (aanzet bij Jeroen, expliciet bij Eva). Tenslotte was nog een zesde manier van waarnemen en verwerken toe te voegen: 6. signalering of/en analyse van de derde verhaallaag (V3), de presentatie door de verteller in het herinnerende "we"-perspectief (Mirjam). In de beide eerste gevallen vindt reductie van de vertellagen plaats, in geval 3 en 4 worden de lagen onderscheiden resp. met elkaar in verband gebracht, en in de gevallen 5 en 6 geven de lezers in metacommentaar blijk van inzicht in de tekststructuur. Een globale verschuiving in de verwerkingsactiviteiten is waarneembaar. Het eerste geval kwam slechts één maal voor, en wel bij de scholieren, het laatste hebben we bij deze tekstpassage ook slechts éénmaal aangetroffen, en wel bij de ervaren lezers. Geval 2 komt relatief vaak bij de scholieren voor, geval 4 is in deze groep echter een uitzondering. Bij de volwassenen en de ervaren lezers zien we duidelijk een verschuiving optreden in de richting van 4: in deze groepen overheerst een contrastbeeld van V1 en V2 gepaard aan een (kritische!) reactie op de gemeenschap. Enkele ervaren lezers illustreren dit met commentaar op de structuur van de tekst (5), wat bij de "doorsnee" volwassenen, hoewel al aanduidingen gemaakt worden, nog minder expliciet gebeurt. De verschuivingen zijn aangegeven in *Figuur 1*. Dit resultaat komt in sterke mate overeen met de bevindingen van Earthman (1992) die in een

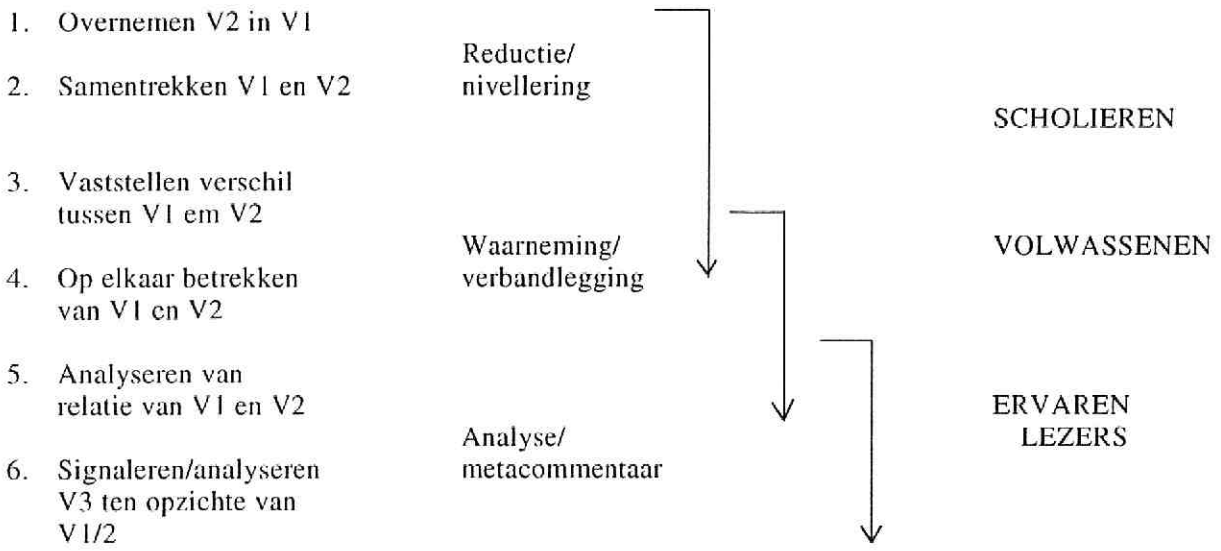


Fig. 1. Niveaus van waarneming en verwerking van de lagen.

vergelijkbare studie eveneens tot de conclusie komt dat minder ervaren lezers "clearly had difficulty in operating at more than one level and in assuming more than one perspective" (p. 378).

3.2.2. Soorten inferenties

Een verwerkingsstrategie die de lezers bij de verwerking van de lagen veelal inzetten, was het maken van inferenties: zij maakten aanvullingen en afleidingen, trokken conclusies en bouwden verwachtingen op. In de meeste gevallen ging het daarbij om het aanbrengen van causale samenhang in de voorstellingen die zij zich vormden. In deze activiteiten manifesteerde zich ten dele de waarneming van de lagen. Er zijn diverse verwerkingsactiviteiten aan te wijzen die aan V1, aan V2 dan wel aan V3 zijn te verbinden. Ik onderscheid hierbij vijf soorten inferenties en conclusies.

1. Veel lezers (re)construeren een intentie, motief of oorzaak voor het handelingsverloop (V1). Zo gissen sommigen dat Emily geen zelfmoord, maar een moord gaat plegen, vermoeden anderen dat ze "echt om Homer geeft", of dat ze "hem niet kan krijgen" omdat hij homosexueel zou zijn.
2. Tevens zijn er lezers die aanvullingen maken voor de gezichtspunten en bedoelingen van de gemeenschap: "ze zijn bezorgd omdat ze omgaat met iemand die niet zo rijk is als zij", "dus het gaat via de dames naar de heren die dus de macht hebben" (V2).
3. Wanneer V1 en V2 contrastief met elkaar in verband gebracht worden, proberen lezers veelal de positie van Emily in de gemeenschap aan te geven hetzij door een karakterisering van Emily ("ze legt gewoon iedereen haar wil op tot de drogist toe"), hetzij door een karakterisering van de gemeenschap ("vooral niet van zou ze ongelukkig zijn of zo, maar nee nee die relatie met Homer kan niet omdat-ie een Noorderling is"). Soms plaatst een lezer ook beide partijen in elkaars licht: "'Arme Emily' zeggen dat is eh, medelijden met haar hebben dat is juist het ergste wat ze zo iemand aan kunnen doen, want andere dingen raken haar niet".
4. Daarnaast zijn er aanvullingen die te maken hebben met het verbinden van informatie uit verschillende delen van de tekst ('bridging inferences'). Bij de hier gebruikte tekstpassages komt dat maar een enkele keer voor. Zo zegt bijvoorbeeld Sean [SCHOL]: "Misschien heeft ze daarmee die ratten vergiftigd die die stank om het huis veroorzaakten". Maar bij de meeste lezers vallen de stukjes pas op hun plaats aan het eind van het verhaal en zelfs dan niet bij iedereen. Het enige soort inferenties dat hier de verbinding naar andere delen van het verhaal

vasthoudt zijn karakteriseringen die een herhaling of generalisering van de situatie of van gedrag inhouden, zoals "dat ze dat rattengif kocht dat wordt als voorbeeld gegeven voor hoe onaantastbaar ze was" (Wil [ERVAR]).

5. Corresponderend met de waarneming van de derde vertellaag houdt een enkele lezer zich bezig met de intenties van de verteller of de auteur – die niet altijd precies gescheiden worden ("Het stuurt allemaal af op de ondergang van die vrouw").

De twee eerste gevallen kunnen in principe op ieder niveau van verwerking van de lagen voorkomen. De gevallen 3 en, in de gevonden vorm, ook 4 zijn veelal manifestaties van de manieren waarop V1 en V2 met elkaar in verband worden gebracht. Het vijfde geval gaat samen met de waarneming van de derde vertellaag.

3.2.3. Vormen van evaluatie

Een tweede categorie reacties waarin de verwerking van de verhaallagen zich neersloeg waren de evaluatieve uitspraken. We hebben twee manieren geregistreerd waarop lezers lieten blijken wat ze vonden. In het ene geval had hun reactie betrekking op inhoudelijke aspecten ("bedompte sfeer", "arrogante vrouw", "zo zie je wat voor een persoonlijkheid ze is", "wat een bemoeizuchtige mensen"). In het andere geval spraken de lezers zich uit over de manier van schrijven of de verhaalstructuur: "het is heel afstandelijk voor me", "ik vind het geen kunst om iemand zo van buitenaf te bekijken". Deze twee soorten evaluaties komen nog duidelijker tot hun recht als de informanten in het interview naar hun oordeel over het gehele verhaal gevraagd wordt. Zij motiveren in hun antwoorden het totale oordeel veelal met verwijzingen naar een van beide mogelijkheden. Ik geef van elk een voorbeeld van een positief en een negatief slotoordeel.

1. Inhoudelijk, positief:
Peggy [SCHOL]:
I: Vind je het een leuk verhaal?
P: Ja
I: Wat vind je er leuk aan?
P: Ik denk de dood van Homer. Echt goed.

2. Inhoudelijk, negatief:
Martin [VOLW]
Ik vind het niet echt een prettig verhaal en het bevalt me niet dat het deprimerend is. Die vrouw lijdt, is eenzaam, ze is psychologisch enigszins instabiel, misschien wel schizofreen, en in ieder geval gedesillusioneerd. Daarom vind ik het geen leuk verhaal, ehm, het was gewoon te deprimerend. En ik mocht die vader niet, ik denk dat ze door zijn schuld zo is geworden. Dus eigenlijk staat het verhaal me helemaal niet aan.

3. Structuur, positief:
Eva [ERVAR]
Tsja, nou merkwaardig, maar ik vond het wel eh vooral hoe dus geschetst werd hoe zo'n dorp reageert op iemand die niet helemaal in hun beeld past, die zich natuurlijk ook afzijdig houdt, maar die, iemand die zich anders gedraagt. ik vind het wel mooi, de stijl was wel mooi, en eh, nou het was vooral ja, indringend, een indringende manier van schrijven.

4. Structuur, negatief:
Jeroen [VOLW]
Ik vind 't ehh, ik vind 't een slap verhaal, ik vind het allemaal heel oppervlakkig blijven. Heel heel afstandelijk beschreven, ehh van een afstand, ik vind het is niks uitgewerkt. Onduidelijke personages, allemaal heel afstandelijk en 't gaat, het is een soort opsomming van.. van een buitenstaander.

Gemiddeld zijn de motiveringen van de scholieren minder uitgewerkt. Antwoorden als "het was saai, "het was niet slecht" zijn zelfs zo algemeen dat niet is uit te maken waarop het oordeel eigenlijk betrekking heeft.[4] Toch verwijzen de paar scholieren die hun oordeel uitvoeriger

toelichten (twee van hen waaronder Shannon) ook naar de manier waarop het verhaal is gestructureerd. Zelfs in het boven aangehaalde antwoord (1) is waarschijnlijk de opbouw van het verhaal mee bedoeld. Enkele duidelijke voorbeelden van beoordeling op grond van sympathie voor of antipathie tegen de Emily, de sfeer, of de buurt zijn met name te vinden bij de volwassen lezers. Bij de meeste volwassen lezers echter speelt ook de verhaalstructuur een rol, zij het dat de oordelen eerder in de negatieve richting wijzen dan bij de ervaren lezers, die veelal gematigd positief of gemengd oordelen. Bij Jeroen speelt waarschijnlijk een rol dat de manier van vertellen een opgaan in het gebeuren en/of identificatie met de hoofdpersoon in de weg staat.

4. DISCUSSIE

Er zijn enkele verschuivingen in de waarneming en verwerking van de gelaagde structuur van Faulkners verhaal zichtbaar geworden. Minder ervaren lezers richten zich eerder op het eigenlijke gebeuren. Meer ervaren lezers vormen zich vaker verschillende voorstellingen: een beeld van de dorpsgemeenschap en hun kijk op de zaak, maar ook een constructie van het perspectief van de hoofdpersoon; daarnaast zijn ze zich sterker bewust van de manier waarop de tekst is opgebouwd. De verschillen manifesteerden zich onder meer in de aanvullingen en afleidingen die de lezers maakten en kwamen ten dele terug in de oordelen over het verhaal. Men ziet hier dus verschillen in het vermogen om *simultaan voorstellingen op te bouwen en die op elkaar te betrekken*. De structuur van het verhaal blijkt dan functioneel: het biedt de lezer als het ware de mogelijkheid door verschillende lenzen naar het verhaalde gebeuren te kijken: Emily in het licht van de gemeenschap, Emily in het licht van haar jeugd, de gemeenschap in het licht van Emily, het geheel in het licht van locatie en geschiedenis. De verschillende voorstellingen zetten aan tot het denken over oorzaken en achtergronden van de handeling en tot een meervoudige kijk op het macabere gebeuren.

Wat zijn nu mogelijke gevolgtrekkingen uit het hier gepresenteerde materiaal? De vergelijking maakt sommige tekortkomingen van minder ervaren lezers zichtbaar en laat tegelijk mogelijkheden voor verbetering of uitbreiding zien. Om met vormen van literaire gelaagdheid en de daaraan verbonden "meervoudigheid" van betekenissen en visies om te kunnen gaan, is de beheersing van een aantal verwerkingsstrategieën en -activiteiten nodig. Samenvattend zijn de volgende te noemen:

1. Het vermogen simultaan voorstellingen op te bouwen, die hetzij in elkaar ingebed zijn, hetzij naast of zelfs tegenover elkaar staan. Het kan gaan om voorstellingen van verschillende visies op eenzelfde gebeuren (visie van de gemeenschap versus Emily's visie),[5] maar ook van bijvoorbeeld een geëngageerde voorstelling (meeleven met Emily) versus een gedistantieerde voorstelling (analytische kijk op de totstandkoming van daad), op een concrete voorstelling van een gebeuren naast een abstract-symbolische (hier bijvoorbeeld de voorstelling van het concrete gebeuren naast een algemenere voorstelling van een individu in een verstikkende, sociale entourage).

2. Verbonden met het opbouwen van simultane voorstellingen is een repertoire van strategieën en activiteiten. We hebben hier geobserveerd:

- Het maken van inferenties zoals causale of motivationele aanvullingen en 'bridging inferences' binnen de verschillende verhaallagen, en met name
- het maken van aanvullingen om verschillende lagen met elkaar in verband te brengen zoals hier bijvoorbeeld door het zoeken naar oorzaken voor Emily's handelen juist in de manier van kijken van de gemeenschap. Een specifieke activiteit is daarbij
- het generaliseren van een beschreven gebeuren of handeling tot een algemene karakterisering van een situatie of personage.
- Het bewust registreren van de eigen reacties en van de manieren waarop de wijze van vertellen daarop van invloed is.

3. Het vermogen een oordeel te vormen dat niet alleen op één van de lagen is gebaseerd, maar op verschillende lagen en vooral ook op de relatie van de verschillende lagen tot elkaar. Van

belang zou tenminste zijn dat lezers zich ervan bewust zijn waarop zij zich precies in hun oordeel richten.

Met de uitbreiding van het repertoire van deze strategieën en handelingen in genoemde opzichten nemen in het algemeen ook omvang en genuanceerdheid van de respons toe. Een volgende stap in het onderwijskundige onderzoek zou kunnen zijn methoden te ontwikkelen en te toetsen om dergelijke vaardigheden gericht te trainen en het repertoire van handelingen te vergroten.

NOTEN

* Het onderzoek waarover hier bericht wordt, is uitgevoerd in samenwerking met Sara Davis (Rosemont College, USA). De meeste ideeën over analyse van zowel tekst als data zijn gemeenschappelijk ontwikkeld (zie Andringa & Davis 1994).

1. Zie voor een overzicht van de opvattingen die en de mate waarin zij vertegenwoordigd zijn in het huidige literatuuronderwijs Janssen (1992).
2. Een heel uitvoerige narratologische analyse van het verhaal inclusief hypothesen over de effecten van de structuur op verwerking en representatie door potentiële lezers is te vinden bij Perry (1979). Het verhaal is voorts gebruikt in een psychoanalytisch georiënteerd lezersonderzoek door Holland (1975).
3. Zie voor een uitvoeriger toelichting op deze methode Andringa (1990) en Andringa & Davis (1994).
4. Grzesik (1983) analyseerde oordelen van scholieren over literaire teksten naar de manier waarop ze waren samengesteld en kwam op grond daarvan tot een reeks van ontwikkelingsniveaus. Ook de oordelen in het onderhavige materiaal zouden volgens deze of nog andere aspecten gerangschikt kunnen worden, maar ik beperk me hier tot de relatie van de oordelen met de verhaalniveaus.
5. Meervoudige perspectieven en visies kunnen door teksten op heel verschillende manieren worden opgeroepen. In 'As I lay dying', een roman van wederom Faulkner, wordt bijvoorbeeld het eigenlijke gebeuren door een aantal verschillende personages verteld. Een vergelijkbare techniek is toegepast door Louis Paul Boon in zijn novelle 'Menuet'. Ook hier is het aan de lezer een representatie te vormen van het eigenlijke gebeuren dat als zodanig niet direct toegankelijk is, en daarbij de voorstellingen van de verschillende personages simultaan in het bewustzijn te houden. Calvino (1991) geeft in zijn lezing over "Meervoudigheid" nog verschillende andere voorbeelden.

LITERATUUR

Andringa, E. (1990). Verbal data on literary understanding. A proposal for protocol analysis on two levels. *Poetics 19,* 231-257.

Andringa, E. & Schram, D. (red.). (1990). *Literatuur in Functie. Empirische Literatuurwetenschap in Didaktisch perspectief.* Houten: Bohn Stafleu Van Loghum.

Andringa, E. & Davis, S. (1994). Literary narrative and mental representation or how readers deal with 'A rose for Emily'. In H. van Oostendorp & R. Zwaan (Eds.). *Naturalistic text comprehension.* Norwood (NJ): Ablex.

Beach, R. (1990). New directions in research on response to literature. In Edmund J. Farrell & James R. Squire (eds.). *Transactions with literature: a fifty-year perspective: for Louise M. Rosenblatt.* N.C.T.E.

Beach, R. & Brown, R. (1987). Discourse conventions and literary inference: Towards a theoretical model. In R.J. Tierney, P.L. Anders, & J. Nichols Mitchell (eds.). *Understanding readers' understanding. Theory and practice* (pp. 147-173). Hillsdale, NJ: Erlbaum.

Calvino, I. (1991). *Zes memo's voor het volgende millennium. De Charles Eliot Norton lezingen 1985-1986.* Amsterdam: Bert Bakker.

Coenen, L. (1992). Literaire competentie: werkbaar kader voor het literatuuronderwijs of nieuw containerbegrip? *Spiegel* 10, 55-78.

Dixon, P., Bortolussi, M., Twilley, L.C., & Leung, A. (1993). Literary processing and interpretation: Towards empirical foundations. *Poetics* 22, 5-33.

Earthman, E.A. (1992). Creating the virtual work: Readers' processes in understanding literary texts. *Research in the Teaching of English,* 26 (4), 351-384.

1634241

Grzesik, J. (1983). Methoden zur Deskription des Textverstehens im Unterricht. Zum Problem der Kategorisierung von Textverstehensleistungen im Verlauf und im Resultat des Unterrichts. *Spiel 2*, 73-99.

Grzesik, J. (1990). *Textverstehen lernen und lehren*. Stuttgart: Klett.

Holland, N.N. (1975). *Five readers reading*. New Haven/ London: Yale UP.

Iser, W. (1975). *Der Akt des Lesens*. München: Fink.

Janssen, Tanja (1992). *Het literatuuronderwijs Nederlands in de bovenbouw van het HAVO en VWO. Resultaten van een nationale enquête*. Amsterdam: Instituut voor de Lerarenopleiding.

Langer, J.A. (1990). The Process of Understanding: Reading for Literary and Informative Purposes. In: *Research in the Teaching of English, 24* (3), 229-260.

Laszlo, J. (1990). Images of social categories vs. images of literary and non-literary objects. *Poetics, 19*, 277-291.

Meutsch, D. & Viehoff, R. (1985). Inferenz- und Elaborationstypen beim literarischen Verstehen von Texten: zum Einfluß von Lese- und Äußerungssituation auf ästhetische und polyvalente Verstehenshandlungen. Siegen: *LUMIS-Schrift*.

de Moor, W. (red.) (1990). *Stiefkind en Bottleneck. De toetsing in het literatuuronderwijs*. Nijmegen.

Perry, M. (1979). Literary dynamics. How the order of a text creates its meanings [With an analysis of Faulkner's 'A rose for Emily']. *Poetics Today 1*, 1-2, 35-64/ 311-361.

Purves, Alan C. (1981). *Reading and literature. American achievement in international perspective*. N.C.T.E.

Schmidt, S.J. (1980/1991). *Grundriß der empirischen Literaturwissenschaft*. Mit einem Nachwort zur Taschenbuchausgabe. Frankfurt am Main: Suhrkamp.

SPIEL (1989). (Siegener Periodicum zur Internationalen Empirischen Literaturwissenschaft) 8, Special Volume: *Developments in literary reading/ Entwicklungen im literarischen Lesen*.

Thomson, J. (1987). *Understanding teenager's reading. Reading processes and the teaching of literature*. Melbourne: Methuen

Willenberg, H. (Hrsg.). (1987). *Zur Psychologie des Literaturunterrichts*. Frankf./M.: Diesterweg.

Winner, E. (1988). *The Point of Words. Children's Understanding of Metaphor and Irony*. Cambridge/ London: Harvard UP.

Zwaan, R. (1993). *Aspects of literary comprehension. A cognitive approach*. Amsterdam/Philadelphia: Benjamins.

Manuscript ontvangen 16-5-1994
Definitieve versie ontvangen 22-11-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 4, pp. 285-299

Beoordeling van het motorische niveau en de schrijfmotoriek van basisschoolleerlingen door leerkrachten

Bouwien C.M. Smits-Engelsman[1], Gerard P. van Galen[1] en Chris G.J. Michels[2]

Katholieke Universiteit Nijmegen,
[1]Nijmeegs Instituut voor Cognitie en Informatie en [2]Onderwijskunde

ABSTRACT

The present investigation was conducted to investigate the prevalence of writing problems in elementary school students (n=734) and to assess the validity of the judgment of student's motor proficiency and writing abilities by their teachers (n=32). From the student sample, 110 children (between 7-12 years of age) were tested on a Test of Motor Proficiency and 48 children were administered writing tasks. The judgement data of the teachers were compared with the measurements on the Motor Proficiency Test and the writing tasks. Based on the teachers ratings the prevalence of severe writing problems was 22%. Students were divided into groups based on teachers' ratings of writing proficiency; these groups differed significantly in their scores on the writing tasks. Teachers' estimation of the handwriting performance was more accurate than of the level of motor proficiency. Cohen's Kappa for the agreement between teachers' judgement data and test data was 0.65 for writing performance and 0.23 for motor performance.

INLEIDING

Lichte motorische stoornissen komen bij tenminste 5-10% van de kinderen in de basisschool leeftijd voor (Gillberg, Gillberg, & Groth, 1989) en worden drie keer zo frequent gevonden bij jongens dan bij meisjes. Deze groep kinderen, met lichte stoornissen in de ontwikkeling van de coördinatie (DCD: Developmental Coordination Disorder), blijkt vooral moeite te hebben met fijne motorische vaardigheden waarbij snelheid en nauwkeurigheid zijn vereist. Het schrijven is een voorbeeld van een vaardigheid die door de complexiteit van de bewegingen een beroep doet op een hoog niveau van coördinatie (Schoemaker, 1992).

In tegenstelling tot de gangbare *produktgerichte* beoordeling van schrijfproblemen willen wij een *procesgerichte* diagnostiek gaan ontwikkelen. In deze laatste aanpak wordt ervan uitgegaan dat schrijfproblemen hun oorsprong vinden in meerdere, verschillend geaarde processen. Aan de hand van een theoretisch model worden verschillende cognitief psychologische processen aangegeven die bij het schrijfproces kunnen zijn gestoord. Zo'n model moet richting geven aan de opsporing van de zwakke schakel(s) binnen het schrijfproces. Hierdoor krijgt men de beschikking over handvatten voor een analytisch onderzoek van de complexe schrijffunctie en zijn afwijkingen.

Volgens het door ons gebruikte model voor handschriftproduktie (Van Galen, 1991) is sprake van een top-down hiërarchie. De hogere, meer cognitieve modules moeten tijdens het schrijfproces eerst zijn doorlopen, voordat de lagere, motorische niveaus kunnen worden bereikt. De output van de hogere cognitieve niveaus dient als input voor het daaropvolgende lagere motorische niveaus. Dit wil niet zeggen dat er sprake is van seriële activiteit. De verschillende modules kunnen tegelijkertijd (parallel) actief zijn met verschillende gedeelten van de te schrijven

Adres: Postbus 9104, 6500 HE Nijmegen.

tekst. In het model van Van Galen worden zes niveaus onderscheiden die betrokken zijn bij handschriftproduktie.

1) Het intentionele niveau
Op dit algemene en meest abstracte niveau van het model wordt de beslissing genomen dat men iets op papier wil zetten.

2) Het linguïstische niveau
Op dit niveau maakt de schrijver gebruik van zijn semantische kennis van de taal en de regels van de grammatica om de concepten te verzamelen die voor het bericht nodig zijn. Hij bedenkt wat hij wil schrijven en maakt daarbij gebruik van opgedane ervaringen.

3) Het lexicale niveau
Hier worden de syntactische en lexicale informatie opgezocht in ons 'geheugenwoordenboek', het mentale lexicon. De woorden moeten worden gevonden en zinnen moeten volgens de juiste grammaticale regels worden geformuleerd.

4) Het foneem-grafeem conversie niveau
Via verschillende lexicale en fonologische bewerkingen moeten de juiste letters en lettersequenties worden bepaald. Hiervoor zijn een goed woordbeeld, kennis en automatisering van spellingsregels noodzakelijk. Tijdens het leerproces zijn auditieve analyse, visuele analyse en synthese, en de intermodale functie erg belangrijk.

5) Het allograaf-selectie en motorprogrammeringsniveau
De keuze voor de feitelijke vormgeving van de letter noemt men de allograafselectie. De allograafselectie bestaat in het model uit twee stappen. Eerst wordt een bepaald lettertype geselecteerd (bv cursief, blokletter, of hoofdletter). Vanuit het lange termijn motorisch geheugen vindt de vertaling van het grafeem naar de allograaf plaats. Deze conversie leidt vervolgens tot de activatie van de aanleerde motorprogramma's, die de ruimtelijke informatie bevatten over de vorm en de volgorde van de opeenvolgende letterhalen (Vergelijk bv de halen in de hoofdletter A en cursief geschreven a).

6) Het executie niveau
Dit niveau bestaat uit twee stappen namelijk het parametrisatie en het initiatie-niveau. Tijdens de parametrisatie vindt de spier onafhankelijke afstemming van het geselecteerde motorprogramma plaats aan de eisen van dat moment qua orientatie, grootte en snelheid. In de initiatie-fase resulteert dit in de recrutering van het juiste type en juiste hoeveelheid motor-units van de spieren die de beweging uitvoeren. De mate van vloeiendheid en nauwkeurigheid waarmee een beweging verloopt wordt bepaald door de afstemming van de verschillende motor-units (kleinste functionele eenheid van de spier). De precisie van de timing wordt bepaald door de afstelling van de samenwerking tussen de verschillende spiergroepen (de agonist en antagonist), de hoeveelheid co-contractie, de mate van oscillatie demping en de hoeveelheid storende meebewegingen.

Het onderzoek naar de ontwikkeling van de schrijfvaardigheid bij kinderen is gericht op het beantwoorden van de volgende vragen: (1) Welke deelprocessen zijn bij kinderen met bepaalde schrijfproblemen verstoord ? (2) In hoeverre verschillen deze problemen van normale ontwikkelingsvertragingen? (3) Kan er een procesgerichte taxonomie worden ontwikkeld die gericht is op de problematiek van deze kinderen. 4) Kan een *procesgerichte* taxonomie dienen als basis voor diagnostische en remediërende oefenschema's in het reguliere onderwijs zodat een verantwoorde foutenanalyse gemaakt kan worden en een theoretisch gefundeerd hulpprogramma ontwikkeld kan worden.

Om enkele van deze vragen te kunnen beantwoorden is een meerjarige onderzoeklijn opgezet. Hierdoor vormt zich geleidelijk een experimenteel en empirisch getoetste domeinspecifieke onderwijstheorie met betrekking tot de motorische aspecten van het schrijven.[1]

In het kader van dit uitgebreide onderzoek bestond de behoefte een beeld te krijgen over het voorkomen (de prevalentie) van schrijfproblemen op Nederlandse basisscholen. Bovendien moest voor het experimenteel onderzoek naar de normale en afwijkende ontwikkeling van het schrijven, het begrip zwakke schrijver worden geoperationaliseerd en moest een aselecte steekproef van de zwakke schrijvers worden getrokken. Het oordeel van de leerkracht over het aanwezig zijn van schrijfproblemen lijkt de meest ecologische variabele die kan worden gekozen. Deze keuze lokt evenwel een validiteitsdiscussie uit; zijn de leerlingen die door de leerkracht als zodanig zijn aangeduid ook inderdaad zwakke schrijvers. Het oordeel van de leerkrachten is van wezenlijk belang omdat zij de eerste lijn vormen voor het signaleren van motorische en schrijfmotorische problemen. Het lijkt daarom aangewezen het oordeel van de leerkracht over de motoriek en schrijfvaardigheid van hun leerlingen aan een nadere analyse te onderwerpen.

De vraag hoe goed leerkrachten motorisch gedrag kunnen observeren en inschatten is in het Verenigd Koninkrijk onderzocht door Henderson en Hall (1982). Deze auteurs lieten de overeenkomst zien tussen het oordeel van leerkrachten en andere manieren om het motorisch presteren te meten. Hieruit bleek dat leerkrachten heel nauwkeurig kunnen zijn in het identificeren van kinderen met motorische afwijkingen in de leeftijdscategorie van 5 tot en met 7 jaar. In andere studies, waarbij is gekeken naar kinderen boven de 7 jaar, werd een lagere overeenkomst gezien (Gubbay, 1975; Keogh, Sugden, Reymand & Clakins, 1979).

Een doel van dit gedeelte van het onderzoek is, inzicht te krijgen in de sensitiviteit en selectiviteit van de signalerende functie van de leerkrachten op het gebied van motorische en schrijfproblemen bij kinderen tussen de 7 en 12 jaar. Hoewel leerkrachten in de dagelijkse praktijk, vooral in het kader van de zorgverbreding, een signalerende taak hebben gekregen, dient men zich te realiseren dat zij hiervoor niet specifiek zijn opgeleid (Kolders & Van de Berg, 1992). Wij vroegen ons af of leerkrachten kinderen met motorische problemen beter zouden kunnen opsporen als zij daarbij gebruik maken van een gestructureerde motorische observatieschaal.

In dit onderzoek worden de volgende vragen beantwoord:

1) Wat is volgens de leerkrachten de prevalentie van motorische en schrijfproblemen bij basisschoolleerlingen (jongens en meisjes) in groep 4, 5 en 6 ?

2a) In hoeverre komt het oordeel van de leerkracht over de motoriek overeen met uitslagen op een motorische test ?

2b) Neemt de overeenstemming tussen het oordeel van de leerkracht over de motoriek en de motorische test toe als zij gebruik maken van een gestructureerde motorische vragenlijst ?

3) In hoeverre komt het oordeel van de leerkracht over het schrijven overeen met uitslagen op visuomotorische testitems en gestandaardiseerde schrijftaken ?

4) Wat zijn volgens de leerkrachten de oorzaken van schrijfproblemen en in hoeverre stemmen die genoemde oorzaken overeen met de gegevens uit de motorische test, visuomotorische testitems en gestandaardiseerde schrijftaken ?

METHODE

Subjecten

Leerkrachten (n=32) van 12 scholen verdeeld over het land hebben over 746 basisschoolleerlingen gegevens geleverd d.m.v. een door ons ontwikkelde Schoolvragenlijst voor schrijfmotoriek met o.a. vragen over schrijfvaardigheid en motoriek. Van een gestratificeerde steekproef van 110 kinderen is voor 98 kinderen een volledige Groninger Motorische Observatieschaal ingevuld en is bij 103 kinderen de motorische test, de Körperkoördinations Test für Kinder (KTK), afgenomen. Bij 48 van deze kinderen zijn het schrijfexperiment en de schrijftaken afgenomen. Voor de zwakke schrijvers (n=24) van deze groep is een lijst met mogelijke oorzaken van schrijfproblemen ingevuld door de leerkracht.

Meetinstrumenten

Schoolvragenlijst voor leerkrachten
De voor dit onderzoek ontwikkelde vragenlijst bevatte 10 vragen en wordt gescoord op een 5-puntsschaal (interne consistentie alpha .93) (zie Tabel 1). Behalve vragen over verschillende aspecten van het geproduceerde handschrift van de kinderen, zijn ook vragen gesteld over de inschatting door de leerkracht van de positie van het kind t.o.v. de norm voor algemene leerprestaties, spelling en algemene motorische vaardigheden. Hierbij werd op de vragenlijst niet gedefinieerd wat volgens de onderzoekers 'de norm' was zodat elke leerkracht zijn eigen impliciete norm voor hetgeen kinderen op een bepaalde leeftijd zouden moeten kunnen, heeft gebruikt.

Groninger Motorische Observatieschaal
De Groninger Motorische Observatieschaal (GMOS) is een checklist voor de motoriek, ontwikkeld door Van Dellen en Kalverboer (1990). Deze checklist moet worden ingevuld door de leerkracht op basis van het dagelijkse gedrag van de te onderzoeken kinderen. De betrouwbaar-

Tabel 1. Schoolvragenlijst voor leerkrachten voor het opsporen van schrijfproblemen.

Naam: Geboorte datum:
Groep: 3 4 5 Geslacht:M/V
Cijfer voor schrijven: 1 2 3 4 5 6 7 8 9 10

1	De vorm van de letters geschreven door dit kind is:					
	goed	O	O	O	O	O slecht
2	De verzorging van het schrijfwerk door dit kind is:					
	slordig	O	O	O	O	O netjes
3	De regelmaat (qua grootte en hellingshoek) van het handschrift van dit kind is:					
	groot	O	O	O	O	O klein
4	De mate van inspanning die het schrijven dit kind kost is:					
	hoog	O	O	O	O	O laag
5	De vloeiendheid (als tegenstelling tot onderbroken of stotende schrift) van het handschrift van dit kind is:					
	hoog	O	O	O	O	O laag
6	Als U dit kind vergelijkt met de eisen die aan leerlingen in deze groep worden gesteld qua schrijven, waar zit dit kind dan t.o.v. de norm:					
	onder	O	O	O	O	O boven
7	Als U dit kind vergelijkt met de eisen die aan leerlingen in deze groep worden gesteld qua spelling, waar zit dit kind dan t.o.v. de norm:					
	boven	O	O	O	O	O onder
8	Als U dit kind vergelijkt met de eisen die aan leerlingen in deze groep worden gesteld qua motorische vaardigheden, waar zit dit kind dan t.o.v. de norm:					
	onder	O	O	O	O	O boven
9	Als U dit kind vergelijkt met de eisen die aan leerlingen in deze groep worden gesteld qua algemene leerprestaties, waar zit dit kind dan t.o.v. de norm:					
	boven	O	O	O	O	O onder
10	Vindt U dat dit kind schrijfproblemen heeft?					
	nee	O	O	O	O	O zeker

heid van de schaal is hoog (interne consistentie alfa >.80). De GMOS bestaat uit 20 vragen waarop gescoord kan worden met een 4-puntsschaal. De scores kunnen worden omgezet in decielen met normscores ingedeeld naar geslacht en leeftijd. De eerste deciel bevat de scores van 20 en de tiende deciel begint, afhankelijk van leeftijd en geslacht, bij 38 (meisjes van 11 jaar) tot 55 punten (jongens van 6 jaar).

De vragen van de GMOS waren in de eerste opzet door de auteurs ingedeeld in drie subcategorieën: het algemeen motorisch functioneren, de fijne motoriek en de grote motoriek. Deze categorieën blijken in verschillende factoranalyses niet op deze wijze te clusteren (Vaessen, 1988). Ook wij hebben maar twee factoren kunnen vinden namelijk een factor grote motoriek en een factor fijne motoriek (zie Tabel 2). Vraag 13 en 14 laadden onvoldoende op beide factoren en zijn in de verdere analyse niet meer gebruikt. Hieruit kan worden geconcludeerd dat de observatieschaal kan worden teruggebracht tot twee onderdelen 1: de 'GMO-grote motoriek' (interne consistentie .95) en 2: de 'GMO- fijne motoriek' (interne consistentie .92). Wij hebben in onze analyses naar de samenhang gebruik gemaakt van de door ons gevonden factoren

Tabel 2. Groningen Motorische Observatieschaal, onderverdeeld in drie subgroepen volgens de oorspronkelijke opgave van de auteurs (van Dellen en Kalverboer, 1990).

Vragen met betrekking tot de fijne motoriek:
3. Schrijft slordiger dan normaal, als ook aan de spelling of inhoud van het geschrevene moet worden gedacht
6. Heeft moeite met activiteiten, die fijne bewegingen vereisen (b.v. handwerken of schrijven)
9. Heeft problemen met de oog – hand coördinatie
12. Schrijft onregelmatiger dan andere kinderen
17. Heeft problemen met knopen vastmaken en veters strikken
19. Is onhandig, laat van alles uit zijn / haar handen vallen

Vragen met betrekking tot de grote motoriek:
2. Heeft moeite met activiteiten waarbij het (hele) lichaam betrokken is (b.v. aankleden en balvangen)
5. Is letterlijk snel uit zijn / haar evenwicht
8. Maakt voor een situatie ' goede ' bewegingen, maar op het verkeerde moment
10. Bewegingen maken een starre, houterige indruk
15. Ziet bij balspelen geen kans tijdig op de naderende bal te reageren
18. Heeft moeite met behendigheidsspelletjes
20. Heeft moeite met het snel (zonder steun) op en aflopen van een trap

Vragen met betrekking tot het algemeen motorisch functioneren:
1. De bewegingen van het kind lijken sterk op die van een jonger kind
4. De bewegingen maken een verbrokkelde indruk; ze zijn niet soepel, er komen onderbrekingen in voor
7. Als het kind met de rechter of linker hand iets doet zie je veel meebewegingen in de andere zijde van het lichaam
11. Heeft moeite met het uitvoeren van ritmische bewegingen
13. Moet nadenken bij bewegingen, die een ander kind automatisch uitvoert
14. Heeft nog geen duidelijke handvoorkeur bij activiteiten als handenarbeid en tekenen
16. Verliest onder spanning heel snel controle over zijn / haar bewegingen

Indeling op 4-puntsschaal:
1 niet van toepassing
2 meer niet dan wel van toepassing
3 meer wel dan niet van toepassing
4 wel van toepassing

De vragen uit de GMOS laden slechts op 2 factoren te weten 'GMO- fijne motoriek' (vraag 3, 4 ,6 en 12) en 'GMO- grote motoriek' (vraag 1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20). Vraag 13 en 14 laden op geen van deze factoren en zijn in de verdere analyse naar samenhang niet gebruikt.

(GMO-grote motoriek en GMO- fijne motoriek) of van de totale lijst zonder vraag 13 en 14, waarnaar wordt gerefereerd als de 'GMO' (interne consistentie .95).

Körperkoördinations Test für Kinder
Körperkoördinations Test für Kinder (KTK) is een test voor de algemene dynamische coördinatie (Kiphard & Schilling, 1974). De KTK bestaat uit vier items: evenwicht op een evenwichtsbalk bij achteruit lopen, hinken over hindernis van schuimrubber blokken, zijwaarts heen en weerspringen over een balkje en zich zijwaarts verplaatsen op twee houten kistjes zonder op de grond te komen. Het doel van het gebruik van de test is het niveau van de algemene dynamische coördinatie vast te stellen en de positie van het kind t.o.v. de normgroep. Het leeftijdsbereik van de test loopt van 5 tot 14 jaar. De normering wordt gegeven per leeftijdtoename van één jaar en er wordt op de items hinken en zijwaarts springen onderscheid gemaakt tussen de prestaties van jongens en meisjes. De ruwe waarden op de test worden omgezet in Motorische Quotiënten met een gemiddelde van 100 en een standaard afwijking van 15. Een kind met een MQ-interval van 100 plus of min een standaarddeviatie wordt als normaal geklasseerd. Kinderen met een MQ kleiner dan 85 worden geklasseerd als 'opvallend', met een MQ kleiner dan 70 als 'gestoord'. De KTK heeft vier normtabelen, namelijk voor normaal ontwikkelde, leergestoorde, gedragsgestoorde en hersenbeschadigde kinderen. De inter- en intrabeoordelaarsbetrouwbaarheid van de test is goed (>.85). De handleiding bij de test en de scoringsformulieren waarborgen een voldoende mate van objectiviteit en standaardisatie bij de uitvoering en scoring van de test. De testen zijn afgenomen door speciaal daarvoor opgeleide fysiotherapeuten.

Visuomotorische coördinatie, schrijftaken en schrijfobservatie
Voorafgaand aan het schrijfexperiment zijn bij de kinderen twee items voor de visuomotorische coördinatie afgenomen, namelijk het kopiëren van twee potloodjes en het trekken van een lijn door een smal straatje (Bruininks, 1978). Bovendien hebben de kinderen per jaargroep een standaard-dictee (Van Dongen & Mommers, 1967) gemaakt. Al deze taken zijn gescoord volgens de testhandleidingen.

Verder is gebruik gemaakt van de experimentele versie van de beknopte beoordelingsmethode voor kinderhandschriften (BHK) (Hamstra-Bletz, De Bie & Den Brinker, 1987). Deze schrijftaak bestaat uit het overschrijven van een standaardtekst (overschrijfkaart van de Landelijke Pedagogische Centra) gedurende vijf minuten of tenminste de eerste vijf regels als het kind erg langzaam schrijft. Het overschrijven gebeurt op ongelinieerd papier. De standaardtekst heeft een oplopende moeilijkheidsgraad bij het vorderen van de tekst. De eerste vijf zinnen hebben het leesniveau van groep drie. Het schrijfproduct wordt beoordeeld aan de hand van dertien dysgrafie-kenmerken. Voor de overschrijfsnelheid wordt het aantal letters dat in vijf minuten wordt geschreven geteld.

Tijdens de afname de schrijftaken is een schrijfobservatie gedaan waarbij wordt gekeken naar de houding, penvatting, algehele motorische onrust (Smits-Engelsman, 1984).

Vragenlijst over oorzaken schrijfproblemen
Door middel van een vragenlijst, waarop mogelijke oorzaken voor schrijfproblemen stonden vermeld, is geïnventariseerd wat volgens de leerkracht de oorzaken waren die het meest van toepassing waren op een bepaald kind (zie Tabel 10).

Procedure
Gegevens voor de analyse van het oordeel van de leerkracht zijn in vier stappen verzameld. Ten eerste is met behulp van een schoolvragenlijst het voorkomen van motorische en schrijfproblemen op de basisschool geïnventariseerd. Aan 32 leerkrachten van 12 scholen, verspreid over Nederland, werd o.a. gevraagd de vragenlijst in te vullen.

Ten tweede zijn de geretourneerde formulieren (n=746) in drie groepen verdeeld. Als indelingsprincipe diende het oordeel van de leerkracht in de vorm van een rapportcijfer voor schrijven. Leerlingen met een zeven of hoger werden als goede schrijver geklasseerd, leerlingen met

een zes als gemiddeld, leerlingen met een vijf of lager als zwak. Uit deze groepen is een gestratificeerde steekproef (N=110) getrokken van minimaal twee tot maximaal vier leerlingen per leerkracht. De steekproef bestond uit 30 zwakke schrijvers, 50 gemiddelde schrijvers en 30 goede schrijvers. De 30 zwakke schrijvers zijn gematcht met 30 goede schrijvers binnen dezelfde school op jaargroep, leeftijd, geslacht en handvoorkeur. Aan de leerkrachten (n=32) is gevraagd een motorische observatieschaal in te vullen over het motorisch gedrag van het kind.

Als derde stap beoordeelden kinderfysiotherapeuten (n=12) de algemene dynamische coördinatie en visuomotorische coördinatie bij deze kinderen met behulp van de KTK (Kiphard & Schilling, 1974) en items uit de Bruininks-Oseretsky (1978). Bovendien werden door een onderwijskundige het standaard-dictee (Van Dongen & Mommers, 1967) en de BHK (Hamstra-Bletz, De Bie & Den Brinker, 1987) afgenomen en de schrijfobservatie (Smits-Engelsman, 1984) uitgevoerd.

Ten slotte werd een experimenteel schrijfonderzoek uitgevoerd door een onderwijskundige en een onderwijspsycholoog, op tien scholen verdeeld over het land. Twee scholen hebben aan het experimentele onderzoek niet verder meegedaan. De totale steekproef voor het experimentele onderzoek bestond uit 48 leerlingen, verdeeld over de vierde, vijfde en zesde jaargroep van het basisonderwijs, waarbij de helft uit goede en de helft uit zwakke schrijvers bestond. Bij deze kinderen (n=48) werd de beoordeling van het schrijven door de leerkracht vergeleken met de beoordeling van het schrift op vorm en kinematische aspecten met behulp van een computerregistratie. Het schrijfexperiment, dat voor elk kind op de eigen school werd afgenomen, duurde per leerling ongeveer 30 minuten. Elke trial begon met de presentatie van een sequentie van vier letters via een computergestuurd beeldscherm. Gedurende de registratietijd (10 seconden) moest de leerling de gepresenteerde stimulus opschrijven. De leerlingen schreven met een speciaal aangepaste (Maarse, Janssen & Dexel, 1988), maar verder normaal schrijvende pen, met een ingebouwd mechanisme voor de registratie van de penkracht. Het schrijfpapier werd bevestigd op een CALCOMP XY- tablet type 2300, zodat het niet kon verschuiven tijdens het schrijven. De totale configuratie werd bestuurd door een IBM PS/2 Model 30 computer. De X- en de Y-positie van de pen, alsmede de uitgeoefende penkracht werden met een frequentie van 100 Hz geregistreerd. De meetnauwkeurigheid van het XY-tablet bedroeg 0.2 mm. De schrijftaken in het experiment belastten drie verschillende facetten van het schrijven: de nauwkeurigheid, de lettergrootte, en het ophalen van het motorisch programma voor de letter. Voor een omschrijving, de data-analyse en resultaten van het schrijfexperiment wordt verwezen naar Smits-Engelsman, Van Galen & Portier (1994).

Voor de totale procedure van het beoordeling van het motorisch niveau en de schrijfmotoriek is het kind ongeveer twee uur onderzocht. Bij de kinderen met een zwakke schrijfprestatie (n=24) is aan de leerkracht gevraagd of de kinderen hulp kregen voor de schrijfproblemen en wat de mogelijke oorzaak zou kunnen zijn. Deze genoemde oorzaken zijn vergeleken met de beschikbare motorische gegevens.

Data-analyse

Allereerst zijn met behulp van beschrijvende statistiek de gegevens van de 746 Schoolvragenlijsten samengevat. De verschillen tussen jongens en meisjes, en tussen links- en rechtshandigen zijn non-parametrisch getoetst. Via een multivariate analyse is gecontroleerd of de gestratificeerde steekproef die heeft deelgenomen aan de motorische test (n=103) en waarbij de GMO (n=98) is afgenomen, een representatieve steekproef vertegenwoordigde van de groepen waaruit ze zijn getrokken. Dit was het geval.

De overeenstemming tussen de Schoolvragenlijst en GMO (door de leerkrachten ingevuld) en de motorische test (KTK) is geanalyseerd met behulp van de Kappa-coëfficiënt. Interacties van jaargroep en schrijfproblemen op de GMO, KTK en de somscore van de Schoolvragenlijst zijn geanalyseerd door middel van ANOVA's. Het significantieniveau voor alle toetsen bepaald op .05.

RESULTATEN

Frequentie van motorische en schrijfproblemen bij de hele steekproef

De onderzochte groep (n=746) bestond voor 50% uit jongens en 50% uit meisjes in de leeftijd van 7.3 — 12.6 jaar (9.2 gemiddeld). De kinderen waren vrij gelijk verdeeld over groep 4 (35%), groep 5 (35%) en groep 6 (30%). De gemiddelde leeftijd in de jaargroepen 4, 5 en 6 was respectievelijk acht (7.4-9.9) negen (7.9-10.9) en tien jaar (8.6-**12.6**). Van de onderzochte kinderen was 12 % linkshandig.

Motoriek

Jongens presteren volgens de leerkrachten significant ($p<0.001$) vaker onder de norm voor motoriek dan meisjes (Tabel 3). Volgens de leerkracht scoort 73% van de jongens op of boven de motorische norm, 27% onder de norm waarvan 6% zeer duidelijk onder de norm. Voor de meisjes liggen deze percentages respectievelijk op 87%, 13% en 2%. Er is geen significante interactie van de positie t.o.v. de motorische norm met links- of rechtshandigheid gevonden.

In de verdere analyse is de groep die 'zeer duidelijk onder de norm' scoort geklasseerd als afwijkend op de 'Motorische norm leerkracht'.

Schrijven

Het gemiddelde cijfer voor schrijven in de onderzochte groep is 6.56 (sd 1.04, range 4-9). Op alle vragen van de Schoolvragenlijst verschillen de scores voor de drie schrijfniveaus (Goede schrijvers :>6; Gemiddelde schrijvers :6 ; Zwakke schrijvers < 6) significant van elkaar . De leerkrachten geven 26% van de jongens en 8% van de meisjes een onvoldoende voor schrijven (< 6).

Voor schrijfproblemen en geslacht is een significante interactie gevonden, in de zin dat over jongens meer schrijfproblemen worden gerapporteerd dan over meisjes (Tabel 3). De leerkrachten gaven bij 32 % van de jongens en bij 11% van de meisjes ernstige schrijfproblemen aan. Het percentage schrijfproblemen verandert niet significant over de jaren. Dit betekent dat de schrijfproblemen niet verdwijnen met langer schrijfonderwijs. Bij linkshandige kinderen is volgens de leerkrachten de vorm van de letters significant vaker slecht, schrijven de kinderen minder vaak vloeiend en komen frequenter schrijfproblemen voor.

Tabel 3. Variabelen uit de Schoolvragenlijst waarop de jongens en meisjes significant verschillen.

Variabelen	Subjecten	Jongens	Meisjes
Aantal	746	369	377
Schrijfcijfer:	6.56	6.25	6.86
4 en 5	17 %	26%	8%
6	28 %	32%	24%
7,8 en 9	56 %	42%	68%
Schrijfproblemen:			
Ernstige	22%	32%	11%
Lichte	36%	37%	35%
Geen problemen	43%	31%	54%
Onder de norm voor:			
Spelling	26%	33%	20%
Leerprestaties	20%	24%	16%
Motoriek	20%	27%	13%
Schrijven	30%	43%	18%

In de verdere analyse hebben we de groep voor wie 'zeer duidelijk schrijfproblemen' werden vermeld, geklasseerd als afwijkend op de 'schrijfnorm leerkracht'.

De positie t.o.v. de motorische norm en het voorkomen van schrijfproblemen vertonen een significante samenhang ($\chi 2$, 32.03, $p<0.001$,). Tussen de positie t.o.v. de motorische norm en de andere vragen over het schrijven bestaat ook een significante maar minder hoge samenhang.

Overeenstemming tussen leerkrachtoordeel de motorische test, visuomotorische testitems en gestandaardiseerde schrijftaken bij de gestratificeerde subset van kinderen

Motoriek op de Groninger Motorische Observatieschaal (GMO)

De resultaten van de GMO-totaal (n=98) zijn samengevat in Tabel 8. De gemiddelde score op de GMO-totaal bedroeg 34 (sd 12). Dat betekent dat de meeste vragen 'niet van toepassing' of 'meer niet dan wel van toepassing' zijn. De range liep van 20-76. Bij de score 20 is geen enkele vraag van toepassing, bij 80 zijn alle vragen van toepassing. Het percentage kinderen dat op of boven het gemiddelde scoort is 60. Van de 98 kinderen scoort 22 % 1 standaard deviatie onder het gemiddelde (score 34-46) en 17% scoort meer dan 1 standaard deviatie onder het gemiddelde (score 47-76). In de verdere analyse is de groep die meer dan 1 standaard deviatie onder het gemiddelde ligt geklasseerd als 'afwijkend op de GMO'.

Overeenstemming tussen motoriek gemeten met de GMO en op de Schoolvragenlijst

De leerkrachten hebben 82% van de kinderen met beide meetinstrumenten hetzelfde geklasseerd (Tabel 4). Bij 5% van de leerlingen gaf de leerkracht aan dat het kind een afwijkende motoriek had terwijl de 'GMO: grote motoriek' het kind als normaal klasseerde. In 13% van de gevallen vond de leerkracht het kind normaal terwijl volgens de 'GMO: grote motoriek' afwijkend was. De overeenstemming tussen beide metingen bedraagt 0.22. De vergelijking van de totale GMO i.p.v. alleen het groot motorische deel van de Observatieschaal verandert weinig aan de resultaten.

Fijne motoriek gemeten met de GMO

Het percentage kinderen dat op of boven het gemiddelde scoort voor het onderdeel fijne motoriek is 55 (score 4-8), 19 % zit 1 standaard deviatie onder het gemiddelde (score 9-12) en 24%

Tabel 4. Motoriekbeoordeling leerkracht versus GMO-totaal en GMO- grote motoriek. Motorische norm 0= normaal 1= afwijkend.

		MOTORISCHE NORM (LK.)		
		0	1	Totaal
GMO- grote motoriek	0	78	5	83
	1	13	4	17
Totaal		91	9	100
K = 0.22				n=98

		MOTORISCHE NORM (LK.)		
		0	1	Totaal
GMO- totaal	0	77	5	82
	1	14	4	18
Totaal		91	9	100
K = 0.20				n=98

Tabel 5. Schrijfbeoordeling leerkracht versus GMO-fijne motoriek. Motorische norm 0= normaal 1= afwijkend. Schrijfproblemen 0= afwezig 1= zeer duidelijk aanwezig.

	SCHRIJFPROBLEMEN (LK.)		
	0	1	Tot
GMO-fijne motoriek 0	40	16	56
GMO-fijne motoriek 1	1	42	44
Totaal K = 0.65	41	58	100 n=98

	SCHRIJFPROBLEMEN (LK.)		
	0	1	Totaal
GMO-totaal 0	38	22	60
GMO-totaal 1	3	37	40
Totaal K = 0.52	41	59	100 n=98

zit meer dan 1 standaard deviatie onder het gemiddelde (score 13-16). In de verdere analyse hebben de groep die meer dan 1 standaard deviatie onder het gemiddelde lag geklasseerd als 'afwijkend op de GMO onderdeel fijne motoriek'.

Overeenstemming tussen fijne motoriek gemeten met de GMO en schrijven op de Schoolvragenlijst

De leerkrachten hebben 82% van de kinderen met beide meetinstrumenten hetzelfde geklasseerd (Tabel 5). Bij één procent van de leerlingen gaf de leerkracht aan dat het kind een normale schrijfmotoriek had terwijl de GMO-fijne motoriek het kind als afwijkend klasseerde en in 16% van de gevallen vond de leerkracht het kind afwijkend terwijl volgens de GMO de fijne motoriek normaal was. De overeenstemming tussen beide metingen bedraagt 0.65. Als de hele GMO wordt gebruikt bedraagt de relatie 0.52.

Motoriek gemeten met de Körper-koördinations Test für Kinder

Het percentage kinderen dat de motorische norm haalt is 63%, 24% wordt als opvallend (meer dan 1 sd onder het gemiddelde: risicogroep) en 13% (>2sd) als afwijkend geklasseerd. Volgens de standaard normale verdeling is de verwachting 84% op of boven de norm, 14% opvallend en 2 % afwijkend. Door de gebruikte stratificatie zijn er meer motorisch zwakke kinderen in de steekproef vertegenwoordigd. In de verdere analyse is de groep die meer dan 2 standaard deviaties onder het gemiddelde lag geklasseerd als afwijkend op de KTK.

Overeenstemming tussen de motorische test en het leerkracht oordeel over motoriek op Schoolvragenlijst

De leerkrachten hebben 86% van de kinderen met beide meetinstrumenten hetzelfde geklasseerd (Tabel 6). Bij 10% van de leerlingen gaf de leerkracht aan dat het kind een normale motoriek had terwijl de test het kind als abnormaal klasseerde. In 4% van de gevallen vond de leerkracht het kind afwijkend terwijl volgens de test de motoriek normaal was. Dit heeft er toe bijgedragen dat de overeenstemming tussen beide metingen uitkomt op Kappa van 0.23.

Tabel 6. Motoriek beoordeling Schoolvragenlijst versus KTK. Motorische norm 0= normaal 1= afwijkend.

		MOTORISCHE NORM (LK.)		
		0	1	Totaal
KTK	0	83	4	87
	1	10	3	13
Totaal		93	7	100
K = 0.23				n=103

Overeenstemming tussen de motorische test en het leerkracht oordeel over motoriek gemeten met de GMO

Het percentage kinderen dat met de motorische test en de GMO-grote motoriek hetzelfde wordt geklasseerd is 80% (Tabel 7). Bij 8% van de leerlingen gaf de leerkracht met de GMO-grote motoriek aan dat het kind een normale motoriek had terwijl de test het kind als abnormaal klasseerde. In 5% van de gevallen vond de leerkracht het kind afwijkend terwijl volgens de test de motoriek normaal was. Dit maakt dat de overeenstemming tussen beide metingen uitkomt op een Kappa van 0.22. Als de GMO-totaal wordt gebruikt verandert er niet veel in het resultaat.

Samenhang tussen de motorische test en het leerkrachtoordeel over schrijven

Op de scores op de Schoolvragenlijst en de motorische test is geen significante relatie gevonden. Er zijn geen significante verschillen aangetoond tussen kinderen met en zonder schrijfproblemen qua motorisch niveau op de gemeten motorische test (Tabel 8).

Verschillen tussen kinderen met en zonder schrijfproblemen

Op de items voor visuomotorische coördinatie bleken de kinderen die volgens de leerkracht schrijfproblemen hebben, significant lager te scoren (Tabel 9). Zij maakten ook significant meer

Tabel 7. Motoriek beoordeling leerkracht op de GMO versus KTK. Motorische norm 0= normaal 1= afwijkend.

		GMO: grote motoriek (LK.)		
		0	1	Totaal
KTK	0	75	12	87
	1	8	5	13
Totaal		83	17	100
K = 0.22				n=98
		GMO (LK.)		
		0	1	Totaal
KTK	0	74	13	87
	1	8	5	13
Totaal		82	18	100
K = 0.20				n=98

Tabel 8. Gemiddelde scores op de GMO (n=98) en KTK (n=103) voor kinderen zonder en met schrijfproblemen. Hogere scores op GMO betekenen slechtere motoriek en hogere scores op de KTK betekenen betere motoriek.

Variabele	Geen Problemen	Wel Problemen	p-waarde
GMO-totaal	26.67	42.85	<0.001
GMO-fijne motoriek	5.98	12.74	<0.001
GMO-grote motoriek	19.40	27.94	<0.001
KTK	87.82	90.37	n.s.

fouten in het dictee en maakten meer spatiële fouten (letter te groot of te klein). De afmetingen van hun letters verschilden niet qua hoogte of breedte en ze hadden na vijf minuten overschrijven ook niet minder letters af. Kinderen met schrijfproblemen kregen een significant lagere beoordeling op de dysgrafieschaal toch waren ze volgens de normering niet 'dysgrafisch'. De leerlingen die volgens de leerkracht schrijfproblemen hebben scoren significant slechter op de items van schrijfobservatie. Zij vertonen een slechtere houding tijdens het schrijven en meer tekenen van algehele motorische onrust. Hun penvatting was niet anders dan van de goede schrijvers.

Oorzaken van schrijfproblemen

In Tabel 10 wordt een overzicht gegeven van de oorzaken van schrijfproblemen zoals door de leerkrachten aangegeven. De leerkrachten noemden slordigheid het meest frequent (91%) als de mogelijke oorzaak voor schrijfproblemen bij de onderzochte kinderen (n=24). Stoornissen in de fijne motoriek scoren 73% als oorzaak. Echter bij 43% van de kinderen geven de leerkrachten

Tabel 9. Gemiddelde scores op visuomotorische testitems en gestandaardiseerde schrijftaken voor kinderen die volgens de leerkracht wel of geen schrijfproblemen hebben (n=48).

Variabelen	Geen problemen	Wel problemen	p-waarde
Schrijfhouding	4.93	9.51	0.001
Penvatting	0.40	0.31	n.s.
Visuomotor test 1	0.07	0.38	0.02
Visuomotor test 2	0.39	1.17	0.001
Dictee	0.90	2.37	0.02
Spatiele fouten	0.73	1.96	0.001
Letterafmeting (mm)	2.7/3.2	2.9/3.2	n.s.
Dysgrafieschaal	2.78	7.60	0.001
Aantal letters per 5 minuten	199	185	n.s.

Tabel 10. Percentage van de kinderen waarbij door de leerkrachten een bepaalde factor wel of niet als oorzaak wordt gezien van schrijfproblemen. Er waren meerdere keuzes per kind mogelijk (n=24).

Oorzaak	Niet	Min of meer	Wel
Faalangst	61	17	21
Slordigheid	0	9	91
Stoornissen in de algemene motoriek	31	26	43
Verminderde algemene leerprestaties	39	26	34
Algemene taalproblemen	30	22	47
Lichamelijke of zintuiglijke tekorten	74	9	17
Stoornissen in de fijne motoriek	9	18	73
Factoren buiten school bv thuissituatie	56	17	23

stoornissen in de algemene motoriek als een mogelijke oorzaak voor de schrijfproblemen. Deze laatste veronderstelling wordt niet statistisch onderbouwd (Tabel 8). Van de zwakke schrijvers krijgt 57% geen extra hulp, 26% systematische hulp en 17% incidenteel hulp.

CONCLUSIE

Over de prevalentie van schrijfproblemen kan samenvattend worden gezegd dat volgens de leerkrachten het schrijven slechts bij 42% van de kinderen helemaal zonder problemen verloopt. Bij 22 % van de leerlingen worden duidelijke schrijfproblemen vermeld. Meisjes hebben beduidend minder schrijfproblemen dan jongens. Problemen met schrijven blijven ook aanwezig in de hogere groepen en er wordt op de onderzochte twaalf scholen weinig systematische hulp verleend aan zwakke schrijvers. Dit hoewel blijkt dat schrijfproblemen niet vanzelf minder worden. Het niet verminderen van de percentages schrijfproblemen in de hogere groepen is in overeenstemming met recente literatuur (Hamstra-Bletz & Blöte, 1993; Mojet, 1991).

De groep kinderen die een onvoldoende voor schrijven haalt (17%) is kleiner dan de groep die volgens de leerkracht problemen heeft. Dit komt waarschijnlijk doordat de impliciete schrijfnorm van de leraar niet bij 6 ligt maar hoger (het gemiddelde lag op 6.56) en doordat sommige kinderen ondanks hun problemen met schrijven toch een redelijk produkt afleveren.

Verder valt op dat van alle onderzochte vaardigheden voor gemiddeld 20% van kinderen een prestatie onder de norm worden aangegeven. Is hier sprake van het zogenaamde Posthumus-effect of zijn de kinderen met een lage score ook objectief gemeten slecht in het schrijven en in hun motoriek? Dit is geverifieerd door de inschatting van de leerkracht te vergelijken met meer objectieve meetinstrumenten voor de motoriek en de schrijfvaardigheid. Wat betreft de inschatting van de leerkracht over de motoriek valt te zeggen dat het door de leerkracht veronderstelde motorische niveau gemiddeld iets hoger ligt dan het niveau gemeten door middel van de motorische test. Volgens de leerkrachten scoort 73% van de leerlingen op of boven de motorische norm, volgens de test maar 63%. De leerkrachten hebben ondanks het feit dat de beoordeling van de algemene motoriek geen deel uitmaakt van hun opleiding (Kolders & Van de Berg, 1992) toch een impliciete norm ontwikkeld over wat het gemiddelde motorische niveau van een kind in een bepaalde leeftijdsgroep zou moeten zijn. De overeenstemming tussen het leerkrachtoor-

deel en de motorische test neemt niet toe als de leerkracht gebruik maakt van een gestructureerde motorische vragenlijst (GMO).

Wat betreft het schrijven kan na analyse van de gegevens worden gesteld dat de leerkrachten met de door ons ontwikkelde Schoolvragenlijst (Tabel 1) schrijfprobleem signaleren die ook worden gemeten door middel van de gestandaardiseerde schrijftaken. De overeenstemming tussen de klassificaties is hoog, zeker gezien het feit dat het zeer verschillende meetinstrumenten betreft. De inschatting van de leerkracht van het niveau van handschriftproduktie komt in hoge mate overeen met de meting door middel van de GMO-fijne motoriek en de gestandaardiseerde schrijftaken. Het is duidelijk dat de norm voor fijne motoriek en schrijven bij de onderzochte leerkrachten verder is genuanceerd dan die voor algemene motoriek. Over de inschatting van het niveau van de schrijfmotoriek valt op dat volgens de leerkracht het niveau gemiddeld hoger zou moeten liggen dan het gemeten niveau met behulp van fijne motorische taken.

Interessant zijn de gegevens over de oorzaken van schrijfproblemen. Leerkrachten leggen zowel in de Schoolvragenlijst als bij het aangeven van oorzaken voor schrijfproblemen een zeer duidelijke relatie tussen algemene motoriek en schrijven. Echter als de motoriek wordt gemeten met behulp van een test voor algemene dynamische coördinatie (KTK) blijkt deze relatie niet aantoonbaar. Kinderen met en zonder schrijfproblemen verschillen motorisch wel van elkaar als het motorische niveau wordt gemeten met een motorische observatieschaal die meer de kwalitatieve aspecten van de motoriek meet.

Ten slotte een opmerking over de meest frequent genoemde oorzaak van schrijfproblemen namelijk slordigheid. Gemeten is dat slechte schrijver meer spatiële fouten maken d.w.z. dat het schrijfspoor van zwakke schrijvers minder vaak tussen de lijnen blijft. Slordigheid wordt door Van Dalen omschreven als onachtzaamheid of zorgeloosheid. Echter de veronderstelling dat dit in 91% van de gevallen de oorzaak is van slecht schrijven moet worden betwijfeld. De kinderen hebben tijdens de testsituatie uiterst gemotiveerd gewerkt maar zij waren blijkbaar toch niet in staat een beter schrijfprodukt af te leveren.

Concluderend kan worden gesteld dat *produktgerichte* beoordeling door de leerkracht een reële inschatting van het niveau van de fijne motoriek en het schrijven geeft. Echter met het leggen van een relatie van het schrijfprobleem met mogelijke onderliggende oorzaken en processen heeft de leerkracht meer problemen. Onderzoek naar het disfunctioneren van de onderliggende processen zal meer helderheid moeten gaan verschaffen over de mogelijke oorzaken van schrijfproblemen. Door diagnostische schrijftaken te ontwikkelen die extra belastend zijn voor een bepaald deelproces van het schrijven kan worden gekeken wat de effecten zijn van die belasting op de bewegingseigenschappen. Zo is bijvoorbeeld gebleken (Schoemaker, 1992) dat kinderen met stoornissen in de algemene motorische coördinatie (houterige kinderen) vaker problemen hebben met het plannen van de beweging (motor programmering) terwijl kinderen met een cerebellaire stoornis tengevolge van aangeboren verminderde schildklierfunctie (congenitale hypothyroidie) vooral moeilijkheden hebben met de uitvoering (executie) van de beweging (Kooistra, 1994). Uit onderzoek van Van Galen et al., (1993) is gebleken dat slechte schrijvers zónder duidelijke algemene motorische coördinatie problemen maar mét stoornissen in de fijne motoriek, met name moeite blijken te hebben met het onder controle brengen van de natuurlijke ruis van het neuromotorische systeem waardoor het handschrift spatiëel onnauwkeurig wordt. In ons onderzoek (Van Galen et al., 1993; Smits-Engelsman et al., 1991, 1993, 1994a,b) wordt een aanzet gegeven tot een, op experimentele grondslag gefundeerde, taxonomie voor schrijfstoornissen in het basisonderwijs. Kennis van de taxonomie, gebaseerd op onderliggende processen, kan de leerkracht helpen om kinderen die behoren tot bepaalde subtypes schrijfproblemen op een meer adequate wijze te begeleiden.

NOOT

1. Dit onderzoek wordt gedeeltelijk gefinancierd door SVO, project 94702, genaamd: Op wetenschappelijke theorie gebaseerde diagnose van de normale en gestoorde ontwikkeling van de schrijfmotoriek.

LITERATUUR

Bruininks, H. (1978). *Bruininks Oseretsky test of motor proficiency.* Minnisota: American Guidance Service.

Gillberg, I.C., Gillberg, C., & Groth, J. (1989). Children with pre-school minor neurodevelopmental disorders. V. Neurodevelopmental Profiles at age 13. *Developmental Medicine and Child Neurology, 31,* 14-25.

Gubbay, S.S. (1975). *The clumsy child.* Philadelphia: Saunders.

Hamstra-Bletz, L. & Blöte, A.W. (1993). A longitudinal study on dysgraphic handwriting in primary school. *Journal of Learning Disabilities, 23,* (10), 689-699.

Hamstra-Bletz, E., De Bie, J., & Den Brinker, B.P.L.M. (1987). *Beknopte beoordelingsmethode voor kinderhandschriften.* Lisse: Swets & Zeitlinger.

Henderson, S.E. & Hall, D.M.B.(1982). Concomitants of clumsiness in young school children. *Developmental Medicine and Child Neurology, 24,* 448-460.

Kiphard, E.J. & Schilling, F. (1974). *Körperkoordinationstest für Kinder.* Weinheim: Beltz Test GmbH.

Keogh, J., Sugden, D., Reymand, C., & Clakins, J.(1979). Identification of clumsy children: Comparison and comments. *Journal of Human Movement Studies,* 5, 32-41.

Kolders, E. & Van de Berg, R. (1992) De leerkracht toetst: een correlationeel onderzoek tussen de Groninger Motorische observatieschaal en de Körperkoordinationstest für Kinder. Afstudeerscriptie Hogeschool West Brabant, afdeling fysiotherapie, Breda.

Kooistra, L. (1994). Psychomotor development in congenital hypothyroidism: Towards a cerebellar hypothesis of CH motor problems. *Journal of Forensic Document Examination, 1.*

Maarse, F.J., Janssen, H.J.J., & Dexel, F. (1988). A special pen for an XY-tablet. In F.J. Maarse, L.J.M. Mulder, W.P.B. Sjouw, & A.E. Akkerman (eds.). *Computers in psychology: Methods, instrumentation and psychodiagnostics* (p. 133-139). Lisse: Swets & Zeitlinger.

Mojet, J.W. (1991). Characteristics of the developing handwriting skill in elementary education. In J. Wann, A.M. Wing, & N. Søvik (eds.). *Development of graphic skills: Research, perspectives and educational implications* (p. 53-75). London: Academic Press.

Schoemaker, M.M. (1992). *Physiotherapy for clumsy children: An effect evaluation study.* PhD Thesis, University of Groningen. Groningen: Stichting Kinderstudies.

Smits-Engelsman,B.C.M. (1984). Begeleiding van het kind met de klinische diagnose MBD/ADD, Deel 1: Het Onderzoek. Hoge School West Brabant.

Smits-Engelsman, B.C.M. & Portier, S.J. (1991). Motoriek en schrijven deel 1: Motorische ontwikkeling en schrijfproblemen. *Nederlands Tijdschrift voor Fysiotherapie, 101,* 235-241.

Smits-Engelsman, B.C.M., Van Galen, G.P., Michels, G.J., & Portier, S.J. (1993). Cross-sectional and longitudinal study on developmental features of psychomotor aspects of handwriting. *Proceedings of the Sixth Handwriting Conference of the International Graphonomics Society: Motor Control of Handwriting* 177-179. Paris: Telecom.

Smits-Engelsman, B.C.M., Van Galen, G.P., & Portier, S.J. (1994a). Poor handwriting in Children: A study on Psychomotor Aspects. *Journal of Forensic Document Examination, 1,* 17-44.

Smits-Engelsman, B.C.M., Van Galen, G.P., & Portier, S.J. (1994b). Cross-sectional and longitudinal study on developmental features of Psychomotor development of handwriting. In Faure, C., Keuss, P., Lorette, G., & Vinter, A. *Advances in handwriting and drawing. A multidisciplinairy approach* (pp. 187-205). Paris: Europia Press.

Vaessen, W. (1988). Ecological validity of hyperactivity studies: Two field experiments with carefully selected hyperactive and clumsy children. PhD Thesis, University of Groningen.

Van Dellen, T. & Kalverboer, A. F. (1990). Groninger Motorische Observatielijst. Laboratorium voor experimentele psychologie, Rijksuniversiteit Groningen.

Van Dongen, D. & Mommers, M.J.C. (1967). Lees- en spellingtoetsen voor leerjaar 1 tot 4 van de lagere school [Reading and spellingtests for grade 1-4 of elementary school]. Nijmegen: Vakgroep Interdisciplinaire Onderwijskunde, University of Nijmegen.

Van Galen, G.P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science, 10,* 165-191.

Van Galen, G.P., Portier, S.J., Smits-Engelsman, B.C.M., & Schomaker, L.R.B. (1993). Neuromotor noise and deviant movement strategies as an explanatory ground for poor handwriting in children. *Acta psychologica, 82,* 161-178.

Manuscript ontvangen 1-9-1994
Geaccepteerd voor publikatie 15-12-1994

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 4, pp. 300-317

Tijdsbesteding in combinatieklassen en enkelvoudige klassen

Een descriptief onderzoek naar besteding van de beschikbare tijd op basisscholen

Marijke Kral[1], Marinus Voeten[2] en Wim Nederlof[1]
[1] *Instituut voor Toegepaste Sociale wetenschappen, Nijmegen*
[2] *Vakgroep Onderwijskunde, Katholieke Universiteit Nijmegen*

ABSTRACT

In this study the amount of allocated time and instructional time are compared for multi-age classes in very small schools and single-age classes in relatively large Dutch primary schools. It is described how much of the available instructional time is spent on the basic subjects (arithmetic, reading and language). In addition, it is investigated how the two types of classes differ in the way teachers organize their instruction during lessons in the three basic subjects. Teachers of single-age (N=36) and multi-age classes (N=92) of grades 2, 4, and 6 kept logs during two weeks. They registered the time actually spent on various subjects and teaching activities. During five lessons on each of the three basic subjects they kept logs of their own activities using a category system. No differences were found between the types of classes in average allocated time and instructional time, nor in the percentages of time actually spent on the basic subjects. As to the latter there were, however, large differences between individual teachers. Students of multi-age classes are more engaged in independent work than students of single-age classes, and they receive less guided exercises and less individual help. Teachers of multi-age classes succeed in spending the available instructional time on average just as efficiently as teachers of single-age classes, but large individual differences exist.

De hoeveelheid tijd die in scholen beschikbaar is om te leren en de kwaliteit van de tijdsbesteding zijn belangrijke variabelen voor de kwaliteit van het onderwijs (Fredrick & Walberg, 1980; Anderson, 1994). In deze studie beschrijven we de verschillen tussen schoolklassen in hoeveelheid daadwerkelijk gerealiseerde instructietijd en in de verdeling van deze tijd over schoolvakken, over instructie-activiteiten van de leerkracht en over groeperingsvormen. Het gaat hierbij met name om verschillen tussen combinatieklassen en enkelvoudige klassen.

De Commissie Evaluatie Basisonderwijs (CEB, 1994) formuleerde onder meer het volgende kwaliteitscriterium: 'De leerlingen krijgen voldoende gelegenheid zich het leer- en vormingsaanbod eigen te maken.' (o.c., p.24). De commissie onderscheidt hieraan twee aspecten. Op de eerste plaats moet er voldoende bruto onderwijstijd zijn om de kerndoelen te bereiken. Het tweede aspect betreft het efficiënt gebruiken van de beschikbare tijd, ofwel het optimaliseren van de onderwijstijd door onnodig tijdverlies te beperken en door een efficiënte klasse-organisatie. De leerkrachten houden zich over het algemeen aan de vastgelegde lestijden, maar volgens de CEB slaagt slechts 63% van de scholen erin om aan alle door de commissie gestelde criteria voor efficiënt tijdgebruik (m.b.t. realisering roostertijd, beperking van lesuitval, reistijd en organisatietijd) te voldoen.

Efficiënte klasse-organisatie is vooral voor combinatiegroepen een belangrijke factor om voldoende onderwijstijd voor alle leerlingen (van alle aanwezige jaargroepen) te realiseren (Lem, Veenman, Nijssen, & Roelofs, 1988). Men zou dan ook kunnen verwachten dat er ver-

[1] Adres: Instituut voor Toegepaste Sociale wetenschappen, Postbus 9048, 6500 KJ Nijmegen.
Deze studie maakt deel uit van het onderzoeksproject *Onderwijs, Instructie en Leerresultaten in Combinatieklassen (SVO-92025).*

schillen bestaan in tijdsbesteding tussen kleine scholen waarin het onderwijs grotendeels in combinatiegroepen wordt gegeven en grote scholen, waarin combinatiegroepen veel minder voorkomen.

Vanuit het perspectief van de leerkracht die onderwijs aanbiedt, kunnen verschillende tijdbegrippen onderscheiden worden (zie o.a. Anderson, 1994): de nominaal beschikbare roostertijd, de feitelijk gerealiseerde roostertijd en de tijd die beschikbaar is voor onderwijs. Het verschil tussen nominale tijd en feitelijk gerealiseerde tijd wordt bepaald door verschijnselen als de schooldag te laat beginnen of langere pauzes nemen dan gepland. De feitelijk gerealiseerde roostertijd wordt niet volledig besteed aan onderwijs; er is bijvoorbeeld tijd nodig voor organisatie van lessen, voor overgangen tussen lessen en om verschillende lokaties te bereiken. Daarnaast wordt tijd besteed aan bijvoorbeeld verjaardagen en andere activiteiten die wel belangrijk kunnen zijn maar die niet direct gericht zijn op de onderwijsinhouden. Verder kan binnen een les tijd 'verloren' gaan aan organisatie (bijv. materialen te voorschijn halen of opbergen) of aan het handhaven van de orde in de klas. Leinhardt (1987) noemt als belangrijke bronnen van tijdverlies externe interrupties tijdens lessen en overgangen tussen activiteiten. De tijd die resteert na aftrek van de tijd die nodig is voor organisatorische en andere niet-onderwijsinhoudelijke activiteiten noemen wij de instructietijd. Dit is de tijd dat leerlingen onderwijs ontvangen in een vak, resp. de tijd dat de leerkracht onderwijs geeft. Het doel van deze studie is het kwantificeren van het onderwijsaanbod voor met name de vakken rekenen, lezen en taal, uitgedrukt in feitelijk voor onderwijs beschikbare tijd. Het gaat in deze studie om de aanbodkant: de gelegenheid die wordt geboden tot leren. De mate waarin de leerlingen deze gelegenheid ook daadwerkelijk benutten (i.c. de taakgerichte leertijd) blijft buiten beschouwing.

Onder onderwijsaanbod verstaan we 'het aanbod van leerinhouden en de wijze waarop de leerstofoverdracht tijdens de lessen wordt georganiseerd' (Harskamp & Suhre, 1992, p. 355). In deze studie komen twee aspecten van het begrip onderwijsaanbod aan de orde. Op de eerste plaats kijken we naar het leerstofaanbod in de vorm van de tijd die beschikbaar is voor onderwijs en meer speciaal naar de feitelijk beschikbare tijd voor de vakken lezen, taal en rekenen. Op de tweede plaats kijken we naar de leerkrachtactiviteiten en de groeperingsvormen binnen de lees-, taal- en rekenlessen. De vraagstellingen voor dit onderzoek zijn:

1. Zijn er verschillen tussen enkelvoudige klassen (in relatief grote basisscholen) en combinatieklassen (in kleine scholen) wat betreft feitelijk gerealiseerde roostertijd en beschikbare instructietijd?
2. Zijn er verschillen tussen enkelvoudige klassen en combinatieklassen in de mate waarin de beschikbare instructietijd wordt besteed aan de basisvakken lezen, taal en rekenen?
3. Hoe verdelen leerkrachten tijdens lees-, taal- en rekenlessen in enkelvoudige klassen en combinatieklassen de beschikbare tijd over diverse instructie-activiteiten en in welke mate worden verschillende groeperingsvormen gehanteerd?

In het onderstaande gaan we eerst in op manieren om het onderwijsaanbod te meten, met name gericht op het meten van de tijd die gemiddeld per week aan een vak wordt besteed. Vervolgens presenteren we gegevens die in de literatuur te vinden zijn over de tijd besteed aan lezen, taal en rekenen in het Nederlandse basisonderwijs. Daarna gaan we nader in op de vraag naar de vormgeving van het onderwijs in lees-, taal- en rekenlessen.

Meten van onderwijsaanbod via logboeken

In onder meer de periodieke peilingsonderzoeken (Wijnstra, 1988; Zwarts, 1990) werd de vragenlijstmethode gebruikt om het onderwijsaanbod te meten. Hetzelfde geldt voor de IEA-onderzoeken (zie bijv. Elley, 1994). Het is echter twijfelachtig of een retrospectieve meting van onderwijsaanbod via eenmalig afgenomen schriftelijke vragenlijsten adequaat is. Een andere methode om onderwijsaanbod te meten is de logboekmethode. Kenmerkend voor deze onderzoeksmethode is dat de gegevens waarin de onderzoeker geïnteresseerd is door de respondenten dagelijks of zelfs onmiddellijk na de betrokken activiteit worden genoteerd.

Harskamp, Suhre & Willemsen (1991) hebben onderzoek gedaan naar de validiteit en be-

trouwbaarheid van de logboekmethode voor het meten van onderwijsaanbod op het gebied van het rekenonderwijs in de basisschool (zie ook Harskamp & Suhre, 1992). Zij maakten daarbij een vergelijking met de meting van het onderwijsaanbod via de vragenlijst die het Cito gebruikte in de periodieke peilingsonderzoeken. Overmaat (1992) verrichtte soortgelijk onderzoek voor het meten van het aanbod van taal en lezen. Een dergelijk onderzoek is ook door Otter (1993) gedaan; zij vergeleek de logboekmethode met de IEA-vragenlijst. Uit deze drie onderzoeken is gebleken dat logboeken ingevuld door leerkrachten redelijk betrouwbare en valide schattingen opleveren voor de tijd besteed aan de vakken rekenen, lezen en taal. Geconcludeerd werd dat de logboekmethode beter geschikt is voor het meten van onderwijsaanbod dan schriftelijke vragenlijsten. Harskamp en Suhre (o.c., p. 362) constateerden dat volgens de vragenlijstgegevens gemiddeld één uur per week meer rekenonderwijs werd gegeven dan volgens de logboekgegevens. De auteurs veronderstellen dat leerkrachten op vragenlijsten de gerealiseerde tijd overschatten omdat ze overwegend de lessentabel weergeven. Otter, Mellenbergh, Blok en Van den Bergh (1992) vergeleken drie methoden voor het meten van de tijd besteed aan taal en lezen. Ze gebruikten zowel een schriftelijke als telefonische vragenlijst, met dezelfde vragen. Verder werd de logboekmethode toegepast. Zij vonden geen statistisch significante verschillen in gemiddelde tijdsbesteding volgens de drie methoden. De correlaties tussen de drie methoden op het niveau van individuele leerkrachten waren echter laag, hetgeen de auteurs tot de opmerking bracht: 'A perfectly valid measure of teaching time seems to be impossible.' (o.c., p. 287). De auteurs bevelen aan meerdere meetmethoden te gebruiken.

Heidelberg, Van den Bergh, Kuhlemeier en Hoeks (1993) richtten zich op leerkrachten in het voortgezet onderwijs en op het vak aardrijkskunde. Zij vergeleken gegevens verkregen via door de leerkracht ingevulde logboeken niet alleen met vragenlijstgegevens maar ook met observatiegegevens. Met betrekking tot de schatting van de gemiddelde tijdsbesteding werden duidelijke methode-effecten aangetoond. Van convergentie tussen de vragenlijsten enerzijds en de logboekgegevens en observaties anderzijds bleek nauwelijks sprake te zijn. De gemiddelden voor de logboekgegevens en de observaties correspondeerden daarentegen redelijk. Ook op het niveau van individuele verschillen tussen leerkrachten bleek de vragenlijstmethode niet te convergeren met de logboekmethode, noch met de observaties, terwijl ten aanzien van de laatste twee methoden wel enige convergentie was waar te nemen.

Op basis van de hier samengevatte onderzoeken concluderen we dat de logboekmethode een betere procedure is om tijdschattingen ten aanzien van onderwijsaanbod te verkrijgen dan de vragenlijstmethode.

Als men voor een logboekmethode kiest, komt onmiddellijk de vraag naar voren, hoe lang de logboeken moeten worden bijgehouden om voldoende betrouwbare gegevens te verkrijgen. De logboekgegevens van Otter et al. (1992) betroffen een aaneengesloten periode van slechts vijf schooldagen. Otter (1993) liet de leerkrachten de logboeken bijhouden gedurende acht weken verspreid over een periode van vijf maanden. Bij Harskamp et al. (1991) ging het eveneens om acht weken gespreid over ongeveer dezelfde periode. Ook bij Overmaat (1992) werden de logboeken acht weken ingevuld, verspreid over 26 weken in de periode oktober-april. Overmaat concludeerde dat acht weken voldoende zijn om een redelijk betrouwbaar beeld te krijgen van de tijdsbesteding aan de meeste leerstofonderdelen binnen de vakken lezen en taal. Ook Harskamp en Suhre (1992) stelden expliciet de vraag aan de orde hoe lang de logboekprocedure zou moeten duren om betrouwbare metingen te verkrijgen. Zij concludeerden dat de variabele 'totale lestijd besteed aan een vak' betrouwbaar te meten is in acht weken of minder (ca. 5 lessen per week); wanneer het gaat om het percentage lestijd besteed aan verschillende leerstofrubrieken binnen een vak te achterhalen, zouden meer weken nodig zijn afhankelijk van de aard van de leerstofrubriek. De genoemde onderzoeken geven geen informatie over de vraag of het beter is om eenzelfde hoeveelheid logboekregistraties uit te smeren over een groot aantal weken dan om het onderwijs gedurende een beperkt aantal weken volledig te registreren. Het laatste zal voor de onderzoeker gewoonlijk de meest praktische optie zijn.

Het onderwijsaanbod voor de vakken lezen, taal en rekenen

Volgens het CEB-rapport (1994) wordt in de groepen 3 t/m 8 32% van de beschikbare tijd besteed aan Nederlandse taal (inclusief lezen) en 18% aan rekenen/wiskunde (o.c., p. 36). Omgerekend in minuten komt dit neer op respectievelijk 480 en 270 minuten. Hierbij gaat het om geplande onderwijstijd volgens de lesroosters. In andere onderzoeken werd de feitelijke gemiddelde tijdsbesteding gepresenteerd op basis van vragenlijsten en/of logboeken. Volgens het peilingsonderzoek van 1990 wordt in de hoogste drie groepen van het basisonderwijs gemiddeld 411 minuten per week taalonderwijs (inclusief lezen) gegeven, terwijl aan rekenen gemiddeld 255 minuten wordt besteed (Zwarts, 1990, p. 35). Er waren geen verschillen tussen de drie jaargroepen. Deze gegevens werden verkregen met de vragenlijstmethode. Overmaat (1992) kwam voor taal in groep 8 op basis van vragenlijsten tot een gemiddelde van 378 minuten (s.d. 91.5) per week en op basis van een logboekperiode van 8 weken tot een veel lager gemiddelde van 270 minuten (geen s.d. vermeld) per week. Harskamp et al. (1991) constateerden dat in groep 8 gemiddeld 216 minuten (s.d. 35.2) per week aan rekenen wordt besteed, gemeten met een 8-weekse logboekperiode. Gemeten met een vragenlijst kwamen zij op een gemiddelde van 282 minuten (s.d. 42.2).

In de hier beschreven onderzoeken werden uiteenlopende gemiddelden gerapporteerd, voor een deel veroorzaakt door het gebruik van verschillende onderzoeksmethoden. Interessanter dan de gemiddelde tijd die aan de onderscheiden vakken wordt besteed, is wellicht de vraag naar verschillen tussen scholen in dit opzicht. De resultaten van Harskamp et al. en Overmaat doen grote verschillen tussen scholen vermoeden. Gezien het verband dat in verschillende onderzoeken wordt gelegd tussen de tijd besteed aan een vak(onderdeel) en de leerprestaties op dat terrein, kunnen die verschillen een bijdrage leveren aan de verklaring van tussen-schoolse prestatieverschillen.

In de 'Beginning Teacher Evaluation Study' (Fisher et al., 1978) die van 1972 tot 1978 in de Verenigde Staten werd uitgevoerd (zie Veenman, Lem, Voeten, Winkelmolen, & Lassche, 1986), werd geconstateerd dat leerkrachten die meer tijd aan een vak of vakonderdeel toekenden, met hun leerlingen betere resultaten behaalden dan leerkrachten die daaraan minder tijd toekenden. In dit onderzoek, waarin overigens ook met logboeken voor leerkrachten werd gewerkt, werden grote verschillen tussen leerkrachten gevonden wat betreft tijdsbesteding. De tijd die in de tweede klas van de basisschool aan wiskundig rekenen werd toegekend, varieerde bijvoorbeeld van 25 tot 60 minuten per dag, hetgeen bleek te leiden tot verschillen in de rekenprestaties. Uit een meta-analyse van Walberg (1986) bleek dat de correlaties tussen op school toegekende tijd en leerprestaties van leerlingen uiteen kunnen lopen van 0.13 tot 0.71 met een mediaan van 0.40. In 95% van de studies werd een positieve samenhang gevonden. Verschillende onderzoekers hebben er overigens op gewezen dat de hoeveelheid toegekende tijd, hoe belangrijk ook, geen voldoende voorwaarde is voor leerprestaties (zie o.a. Gettinger, 1989). Voldoende beschikbare onderwijstijd is echter wel een noodzakelijke voorwaarde voor leren.

Klasse-organisatie en leerkrachtactiviteiten tijdens lees-, taal- en rekenlessen

Efficiënte besteding van de onderwijstijd hangt niet alleen samen met zaken als het realiseren van de roostertijden en het beperken van lesuitval maar ook met een goede klasse-organisatie en een effectieve vormgeving van de instructie.

Onder klasse-organisatie verstaan we 'het treffen van maatregelen en voorzieningen die nodig zijn om een omgeving in het leven te roepen en in stand te houden waarin instructie en leren kunnen gedijen' (Veenman, Lem, Roelofs, & Nijssen, 1992). Meer concreet gaat het om zaken als het klaarleggen en uitdelen van materialen, het organiseren van lesovergangen, het indelen van het lokaal, het maken van regels en afspraken, ordehandhaving en dergelijke. Naarmate een leerkracht de klas c.q. de lessen beter organiseert zal binnen de les minder tijd verloren gaan aan organisatorische zaken en dus meer tijd voor onderwijs beschikbaar zijn. Het CEB-rapport (1994) vat resultaten samen van een onderzoek naar tijdsbesteding tijdens rekenlessen. Geconstateerd werd dat 85% van de leraren niet meer dan 10% van de rekenles nodig had

voor klasse-organisatie. Anderen komen op een hoger percentage; Roelofs (1993, p. 130) bijvoorbeeld constateerde dat gemiddeld 15% van de lestijd opgaat aan organisatorische zaken en dat er leerkrachten zijn die meer dan 25% van de lestijd besteden aan lesovergangen. In combinatieklassen is vanwege de aanwezigheid van meerdere jaargroepen, sprake van een grotere organisatorische complexiteit. Hierdoor staat de leertijd extra onder druk (Veenman, 1987). Alleen al het wisselen tussen de (jaar)groepen kan de benodigde organisatietijd vergroten. De manier waarop leerkrachten omgaan met overgangen tussen activiteiten is van groot belang voor een effectieve tijdsbesteding door de leerlingen (Arlin, 1979).

Onderzoek naar de relatie tussen leerkrachtgedrag en de leerprestaties van de leerlingen heeft een aantal effectief gebleken instructiegedragingen opgeleverd. Rosenshine & Stevens (1986) vatten de studies naar effectieve instructie samen en kwamen tot zes onderwijsfuncties: terugblik, presentatie, inoefening, terugkoppeling, individuele verwerking en periodieke terugblik. Belangrijk hierbij is voor dat elk van deze instructiefasen voldoende tijd wordt uitgetrokken. In combinatieklassen zijn leerlingen, in vergelijking met leerlingen in enkelvoudige groepen, vaak langdurig zelfstandig, zonder directe begeleiding van de leerkracht, aan het werk. Lem et al. (1988) stellen dat in gecombineerde groepen de uitleg, demonstratie en geleide oefening in veel gevallen in krappe tijd aan de orde worden gesteld, terwijl juist deze leerkrachtactiviteiten noodzakelijk zijn voor een zinvolle zelfstandige verwerking.

Veenman et al. (1986) constateerden op basis van een observatiestudie dat in de combinatieklas de taakgerichte leertijd van de leerlingen gemiddeld lager lag dan in de enkelvoudige klas (72% om 78%). De grootste verschillen in taakgerichte leertijd tussen klassen kwamen voor in de groep van combinatieklassen (spreidingsbreedte 45%-89%). In de combinatieklas verwerkten de leerlingen de leerstof significant meer individueel dan in de enkelvoudige klas. De taakgerichte leertijd van de leerlingen bleek mede in relatie te staan tot het didactisch handelen van de leraar. Wanneer de leraar een groter deel van de lestijd besteedde aan klassikale instructie, nam de taakgerichte leertijd toe; bij procedureel gedrag en zelfstandig werken daarentegen af. Het bleek dat het klasse-management van leraren een belangrijke rol speelde bij het realiseren van taakgericht gedrag van de leerlingen. Niet alleen de leerkrachtactiviteiten spelen dus een rol bij de taakgerichte leertijd maar ook de setting waarin gewerkt wordt: bijvoorbeeld klassikaal of individueel.

METHODE

Onderzoeksgroep

Het onderzoek werd uitgevoerd bij een landelijke steekproef van 34 kleine en 15 grote Nederlandse basisscholen. Met een kleine school bedoelen we een school met minder dan 110 leerlingen. In een dergelijke school zullen de leerlingen hun hele basisschoolloopbaan in een combinatiegroep doorbrengen. Onder een grote school verstaan we een school met 250 of meer leerlingen. Een dergelijke school kent overwegend enkelvoudige klassen. Er werd een gestratificeerde aselecte steekproef getrokken. Als stratificatiecriteria werden gebruikt schoolgrootte en gemiddeld leerlinggewicht. Voor een uitvoerige bespreking van de steekproef zie Kral (in voorbereiding).

Aan de leerkrachten van de groepen 4, 6 en 8 (N=144) werd gevraagd gedurende twee weken een logboek bij de houden. In het geval van combinatiegroepen had het logboek alleen betrekking op de bedoelde groep (4, 6 of 8). Uiteindelijk konden wij beschikken over bruikbare gegevens van leerkrachten van 128 groepen (92 combinatiegroepen en 36 enkelvoudige groepen).

Logboek

Het logboek had als doel de registratie door de leerkrachten zelf van (1) de tijd die zij feitelijk besteden aan de verschillende vakken of leeractiviteiten, (2) de besteding van de instructietijd tijdens lees-, taal- en rekenlessen, en (3) het percentage taakgerichte leerlingen op enkele vooraf

bepaalde momenten in de les. Het laatste blijft in deze studie buiten beschouwing, omdat het met de gevolgde logboekprocedure niet op een betrouwbare wijze kon worden vastgesteld (Wisse, 1993; Nederlof, 1994).

Het logboek bestond uit twee gedeelten: een weekoverzicht en lesoverzichten. Het *weekoverzicht* had naast een voorblad waarop een aantal algemene gegevens werden genoteerd (leergroep, datum, roostertijden) vijf dagoverzichten waarop per dag begin- en eindtijden van lessen en andere activiteiten en de aard van de betreffende les of activiteit werden bijgehouden. Ook pauzes en tijd nodig voor organisatie (indien méér dan een minuut) dienden te worden genoteerd. De leerkrachten beschikten hierbij over een categorieënsysteem.

Voor de vakken lezen, taal en rekenen werd de tijdsbesteding gedetailleerder bekeken door middel van het bijhouden van *lesoverzichten*. Daarmee werd informatie verzameld over (de tijdsduur van) leerkrachtactiviteiten en over de (sub)groep waarop de leerkracht tijdens deze activiteiten gericht was (doelgroep). De categorieën waarin gescoord moest worden, zijn gebaseerd op observatie-onderzoek naar het omgaan met combinatieklassen (Veenman et al., 1986; Lem et al., 1988; Roelofs, 1993). Ter illustratie geeft Figuur 1 een voorbeeld van een scoreformulier. De leerkrachten dienden het formulier in te vullen tijdens het lesgeven. Telkens wanneer sprake was van een wisseling van activiteit moest het begintijdstip worden genoteerd; bovendien moest dan worden aangegeven op welke doelgroep de activiteit was gericht. Uit onderzoek van Wisse (1993) was gebleken dat deze procedure voor de leerkrachten hanteerbaar was.

De leerkrachten beschikten over een uitgebreide handleiding waarin de logboekcategorieën en de procedure werden toegelicht. Voor een beschrijving van de ontwikkeling van het logboek verwijzen we naar Wisse (1993); een uitvoerige uiteenzetting van het logboek en de gevolgde procedure is te vinden in Nederlof (1994). In de resultatensectie worden de diverse categorieën beknopt omschreven.

Mede om de leerkrachten niet te veel te belasten, werd gekozen voor een logboekperiode van slechts twee weken. Dit is een korte periode in vergelijking met de studies van Harskamp et al. (1991) en Overmaat (1992). De roostertijden en feitelijke les-, pauze- en organisatietijden

		*** begintijd: ..... uur**						*** eindtijd: ..... uur**					
	**** leerkracht***	*in hokjes precieze tijdstip aangeven*											
1	terugblik												
2	instructie												
3	begeleid (in)oefenen												
4	individuele hulp												
5	nabespreking												
6	herhaling												
7	andere jaargroep												
8	organisatie												
9	pauze												
na afloop codes invullen:													
	*** doelgroep**	*'X' in betreffend hokje zetten*											
1	klassikaal (allen)												
2	hele jaargroep 4/6/8												
3	subgroep binnen												
4	subgroep tussen												
5	andere jaargroep												
6	geen contact												

Fig. 1. Formulier voor het lesoverzicht in combinatiegroepen.

werden gedurende de totale periode geregistreerd. We gaan er vanuit dat twee weken voldoende zijn om een redelijk betrouwbare schatting te verkrijgen van de mate waarin gemiddeld van de roostertijden wordt afgeweken ('verloren' tijd en organisatietijd). Van verschijnselen als lesuitval en excursies zal in deze korte periode geen goed beeld worden verkregen. Volgens de gegevens van de CEB (1994) blijkt lesuitval overigens niet zoveel voor te komen. Slechts bij 8% van de scholen was er gedurende een bepaalde maand sprake van lesuitval. Wij streefden naar logboekregistraties gedurende twee weken waarin weinig of geen bijzondere zaken aan de hand zijn. Dit betekent dat de weektotalen waarschijnlijk een overschatting te zien zullen geven als men ze naar een heel schooljaar extrapoleert.

De tijdsbesteding binnen lessen werd voor vijf lessen per vak (in het totaal 15 lessen) verspreid over de twee weken geregistreerd. Het ging hierbij niet om de tijd besteed aan verschillende lesinhouden, zoals in de onderzoeken van Harskamp et al. (1991) en Overmaat (1992), maar om de verdeling van de tijd over instructie-activiteiten en groeperingsvormen. Deze hangen samen met de (betrekkelijk stabiele) instructiestijl van de leerkracht. We nemen dan ook aan dat voor dit doel met minder lessen volstaan kan worden dan voor het onderzoek naar lesinhouden nodig is. In de observatiestudies naar het onderwijs in combinatieklassen van Veenman et al. (1986), Lem et al. (1988) en Roelofs (1993) werden resp. 20, 8 en 8 lessen per leerkracht geobserveerd om een beeld te krijgen van de tijdsbesteding binnen lessen.

Hoewel wij geen uitspraken wilden doen over de verdeling van de tijd over de verschillende lesinhouden, hebben wij wel beknopte informatie hierover verzameld om een beeld te krijgen van de variatie in lesinhouden. Instructie-activiteiten en groeperingsvormen kunnen immers samenhangen met de leerstof. De geregistreerde lessen gaven een grote spreiding in leerstofonderdelen te zien. Technisch lezen kwam in 52% van de leeslessen aan bod, begrijpend lezen in 34% van de lessen en recreatief lezen ten slotte in 20%. In 55% van de taallessen was spelling aan de orde; ontleden/woordbenoeming, schrijven en spreken/luisteren elk in zo'n 17%. In 59% van de rekenlessen stonden de basisvaardigheden centraal, cijferen in 44%, verhoudingsrekenen en meten/tijdrekenen elk in ongeveer een kwart van de lessen. De verdeling naar lesinhoud verschilde nauwelijks tussen combinatiegroepen en enkelvoudige groepen. Wel waren er wat voor de hand liggende verschillen tussen de drie jaargroepen. Technisch lezen bijvoorbeeld kwamen we vooral in groep 4 tegen en begrijpend lezen vooral in groep 8. Met het oplopen van de jaargroep kwamen we vaker ontleden en minder vaak spelling tegen.

Betrouwbaarheid van het logboek

Ten aanzien van het weekoverzicht geven de boven vermelde studies van Harskamp en Suhre (1992) en Overmaat (1992) voldoende indicaties voor de betrouwbaarheid van de gegevens. De betrouwbaarheid van de lesoverzichten is nagegaan door Wisse (1993). Zij vergeleek logboekgegevens met observaties. In 18 lessen (lezen, taal en rekenen) van zes leerkrachten in combinatiegroepen hield de leerkracht het lesoverzicht bij, terwijl dit tegelijkertijd ook werd gedaan door een in de klas aanwezige observator. Over alle lessen heen bleken de verschillen tussen leerkracht en observator in gemiddelde tijd per categorie klein te zijn. De correlaties varieerden per categorie tussen 0.88 en 0.99. Geconcludeerd werd dat de logboekprocedure voor wat de registratie van de leerkrachtactiviteiten en de doelgroep betreft, voldoende betrouwbaar was.

Procedures

De logboekgegevens werden verzameld in de periode van begin maart tot en met eind april 1993. Aan leerkrachten werd gevraagd om binnen deze periode het logboek gedurende twee weken bij te houden. Dit hoefden geen aaneengesloten weken te zijn. Belangrijk was wel dat de leerkrachten zo 'gewoon mogelijke' weken kozen, dat wil zeggen weken zonder vrije dagen, excursies, festiviteiten en dergelijke.

De leerkrachten werd gevraagd om in de eerste week op maandag, woensdag en vrijdag lesoverzichten bij te houden, en in de tweede week op dinsdag en donderdag. In principe ging het om de eerste lees-, taal- en rekenles van de dag. Per vak diende het logboek voor vijf lessen te worden bijgehouden. Niet alle leerkrachten hebben vijf lesoverzichten ingevuld; deze zijn

toch in de analyse meegenomen. Om de vergelijkbaarheid van de gegevens voor de verschillende leerkrachten te vergroten, zijn in een aantal gevallen (bijv. excursies) ontbrekende gegevens aangevuld (zie Nederlof, 1994). Dit heeft tot gevolg dat de weektotalen voor de instructietijd besteed aan lezen, taal en rekenen op jaarbasis waarschijnlijk als een overschatting moeten worden gezien.

RESULTATEN

De eerste en tweede vraagstelling kunnen beantwoord worden aan de hand van de weekoverzichten. Het gaat hier om een kwantificering van het totale onderwijsaanbod en van het aanbod voor de vakken lezen, taal en rekenen, uitgedrukt in gemiddelde tijd besteed per week. Voor de beantwoording van de derde vraagstelling wordt een beroep gedaan op de lesoverzichten. Hierbij worden de gegevens uitgedrukt in gemiddeld percentage tijd besteed per les. Alle analyses zijn uitgevoerd op het niveau van de leerkracht, c.q. groep. Er is steeds univariaat getoetst, op 5-procentsniveau.

Roostertijd en feitelijk gerealiseerde tijd

Tabel 1 bevat de gemiddelden van de roostertijden en van de feitelijk gerealiseerde lestijd per week. Bij de totale roostertijd per week zijn pauzetijden niet meegeteld. De groepen 6 en 8 kwamen gemiddeld op ongeveer 24.5 uur per week en groep 4 op 22 uur per week. Er waren in dit opzicht geen systematische verschillen tussen enkelvoudige groepen en combinatiegroepen.

Vervolgens keken we naar feitelijke begin- en eindtijden in vergelijking met roostertijden. Tabel 1 geeft de gerealiseerde lestijd (B) als percentage van de roostertijd (A). Zowel enkelvoudige groepen als combinatiegroepen bleken de beschikbare roostertijd ook daadwerkelijk te realiseren (gemiddeld over de jaargroepen 98%). Per dag gingen gemiddeld ongeveer 5 à 6 minuten verloren als gevolg van te laat beginnen of te vroeg ophouden. In een kwart van de gevallen liep het tijdverlies op tot 10 à 15 minuten per dag (niet in tabel). Deze gegevens komen redelijk overeen met die in het CEB-rapport (1994, p. 37).

Hoeveel van de feitelijk gerealiseerde lestijd kan beschouwd worden als daadwerkelijk op de inhoud van het onderwijs gerichte lestijd (instructietijd)? Om dit na te gaan werden in de weekoverzichten van het logboek de categorieën 'organisatie' en 'overige activiteiten' opgenomen. 'Organisatie' werd gedefinieerd als tijd die nodig is om het onderwijs goed te laten verlopen, maar die niet op de onderwijsinhoud is gericht. Voorbeelden zijn tijd die nodig is om naar de gymzaal of het zwembad te gaan en tijd voor het pakken van materialen of het klaarzetten van apparatuur. Met 'overige activiteiten' werd bedoeld activiteiten die los staan van de onderwijsinhoud, zoals een verjaardag vieren of melk drinken. Deze twee categorieën omvatten niet alleen periodes tussen lessen maar ook activiteiten tijdens lessen. Overigens hoefden de leerkrachten activiteiten die minder dan een minuut duurden niet te registreren. In Tabel 1 is de instructietijd uitgedrukt als een percentage van de gerealiseerde lestijd. Gemiddeld bleek sprake te zijn van een reductie van de tijd met ongeveer 15%. Als gevolg van organisatie alleen, ging gemiddeld ongeveer 10% van de tijd 'verloren' (niet in tabel).

Leerkrachten van combinatiegroepen rapporteerden gemiddeld minder tijd voor organisatorische en overige activiteiten dan leerkrachten van enkelvoudige klassen, terwijl we het omgekeerde zouden verwachten vanwege het feit dat de combinatieklas in organisatorisch opzicht complexer is dan de enkelvoudige klas. Alleen in jaargroep 6 was het verschil statistisch significant. Voor deze jaargroep bleken de leerkrachten van enkelvoudige groepen (in grote scholen) gemiddeld 50 minuten per week meer aan organisatie te besteden dan de leerkrachten van combinatiegroepen (in kleine scholen).

In de hogere groepen werd een groter percentage van de lestijd daadwerkelijk aan de onderwijsinhouden besteed dan in lagere groepen. Er waren aanzienlijke verschillen tussen leerkrachten in instructietijd, variërend van 67% tot 100% (niet in de tabel). Een belangrijke factor bij deze verschillen is de tijd die groepen nodig hebben om de gymzaal of het zwembad te bereiken.

Tabel 1. Lestijd (in minuten per week) volgens het rooster, percentage feitelijk gerealiseerde lestijd en het percentage van de feitelijk gerealiseerde tijd dat inhoudelijk wordt besteed (=instructietijd).

	Enkelvoudige groepen		*Combinatiegroepen*		
	gem.	sd.	gem.	sd.	t
Jaargroep 4 :	(n=13)		(n=30)		
- lestijd volgens rooster (A)	1311.9	31.3	1331.0	38.3	-1.58
- % gerealiseerde lestijd (B/A)	98.3	2.0	97.9	2.6	.54
- % instructietijd (C/B)	82.6	8.1	83.7	5.2	- .55
Jaargroep 6 :	(n=10)		(n=32)		
- lestijd volgens rooster (A)	1471.5	13.1	1462.2	19.9	1.38
- % gerealiseerde lestijd (B/A)	98.2	1.7	97.9	2.2	.30
- % instructietijd (C/B)	84.1	6.2	87.9	4.6	-2.08 *
Jaargroep 8 :	(n=13)		(n=30)		
- lestijd volgens rooster (A)	1465.4	12.8	1465.5	12.3	- .03
- % gerealiseerde lestijd (B/A)	98.0	1.8	97.8	2.1	.29
- % instructietijd (C/B)	87.7	3.4	89.1	4.7	- .91

A = roostertijd, B= gerealiseerde lestijd en C = instructietijd.
* $p < .05$ (tweezijdig)

Sommige scholen verliezen per week nauwelijks lestijd aan de organisatie van de gym- en de zwemlessen. Voor andere scholen bleek deze tijd op te kunnen lopen tot 1 1/2 uur of meer per week.

Tijd besteed aan de vakken lezen, taal en rekenen

De totale tijd besteed aan een vak werd berekend door alle inhoudelijk op het vak gerichte activiteiten per week (zoals ingevuld in de weekoverzichten) te sommeren en vervolgens de weektotalen te middelen. Organisatietijd en tijd voor overige activiteiten tijdens de lessen werden dus niet meegerekend. Tabel 2 bevat de gemiddelde tijd die per week werd besteed aan lezen, taal en rekenen apart en voor de drie vakken samen. De tijd is weergegeven in aantal minuten per week en als een percentage van de instructietijd (de gerealiseerde lestijd exclusief 'organisatie' en 'overige activiteiten').

In groep 4 registreerden we in de totale steekproef gemiddeld 220 minuten voor rekenen, in groep 6 was dat 246 minuten en in groep 8 weer wat minder, namelijk 228 minuten. Dit laatste ligt niet veel hoger dan wat Harskamp et al. (1991) vonden voor groep 8; zij kwamen op 216 minuten per week. De standaardafwijkingen zijn vrij groot (in de orde van 20% van het gemiddelde). Voor de middelste 50% van de frequentieverdeling kan het verschil in tijd besteed aan rekenen oplopen tot ongeveer 1 uur per week (niet in tabel).

Uitgedrukt als een percentage ten opzichte van de totale instructietijd, werd gemiddeld ongeveer 20% van de tijd aan rekenen besteed. Alleen in de groepen 8 van combinatiegroepen vonden we een wat lager getal, nl. 17%. Zoals al vermeld, geeft het CEB-rapport (1994) te zien dat aan rekenen 18% van de bruto onderwijstijd wordt besteed, dat is 270 minuten per week.

Aan lezen werd in de groepen 4, 6 en 8 respectievelijk gemiddeld 117, 129 en 115 minuten besteed. Voor taal zijn deze getallen 200, 207 en 215 minuten. Overmaat (1992) kwam voor taal en lezen samen voor groep 8 op een gemiddelde van 270 minuten per week, terwijl wij een veel hoger gemiddelde van 330 minuten noteerden. Aan lezen werd in de groepen 4 en 6 gemiddeld 10 à 11% van de instructietijd besteed, in groep 8 was dat 8 à 9%. Aan taal werd in groep 4 gemiddeld zo'n 18% van de instructietijd besteed, in groep 6 en 8 rond 16%. Kwamen we voor rekenen op een wat hoger percentage uit, voor taal en lezen samen vonden we dus een lager percentage dan de CEB, die op 32% van de bruto onderwijstijd, zijnde 480 minuten per week, uitkwam.

Ondanks twee statistisch significante verschillen tussen combinatieklassen en enkelvoudige klassen, kunnen we over de hele linie concluderen dat er geen systematische verschillen waren in instructietijd voor de basisvakken tussen deze twee typen klassen. Wel was er een aanzienlijke spreiding tussen leerkrachten. Voor taal lagen de standaardafwijkingen evenals bij rekenen rond 20% van de gemiddelden, voor lezen lagen de standaardafwijkingen zelfs boven 30% van de gemiddelden. De middenspreiding (=het verschil tussen het 75e en 25e percentiel, niet in tabel) lag bij taal en lezen net als bij rekenen rond 60 minuten op weekbasis.

In groep 4 werd gemiddeld ongeveer de helft van de instructietijd besteed aan de drie basisvakken rekenen, taal en lezen; in groep 6 was dat ongeveer 47% en in groep 8 ongeveer 44%. Uitgedrukt in uren per week varieerde de middenspreiding van bijna 1 uur (enkelvoudige groepen 4) tot meer dan 2 1/2 uur (combinatiegroepen 8). Voor de groepen 4 en 8 was er een aanzienlijk verschil in spreiding tussen enkelvoudige groepen en combinatiegroepen. De standaardafwijking in de combinatiegroepen was ongeveer het dubbele van die in de enkelvoudige groepen. Dit verschil zal verband houden met het gegeven dat in combinatiegroepen de tijd besteed aan rekenen en taal positief gecorreleerd waren. Het is niet zo, dat meer tijd besteden aan rekenen ten koste gaat van de tijd voor taal of lezen, of omgekeerd.

Tabel 2. De gemiddelde tijd per week feitelijk besteed aan rekenen, lezen en taal in minuten en als percentage van de instructietijd (C).

	Enkelvoudige groepen		*Combinatiegroepen*		
	gem.	sd.	gem.	sd.	t
Jaargroep 4 :	(n=13)		(n=30)		
- feitelijke tijd rekenen-taal-lezen	523.6	58.3	541.6	105.9	- .57
. % van de totale instructietijd (C)	49.4	6.3	49.7	7.7	- .11
- feitelijke tijd rekenen	210.7	43.7	223.8	53.6	- .78
. % van de totale instructietijd (C)	19.8	4.0	20.5	4.4	- .51
- feitelijke tijd lezen	119.5	38.2	116.2	41.6	.25
. % van de totale instructietijd (C)	11.3	3.4	10.6	3.5	.58
- feitelijke tijd taal	194.0	33.3	202.2	51.2	- .53
. % van de totale instructietijd (C)	18.4	4.1	18.6	4.0	- .16
Jaargroep 6 :	(n=10)		(n=32)		
- feitelijke tijd rekenen-taal-lezen	571.8	70.9	584.4	73.5	- .48
. % van de totale instructietijd (C)	47.2	6.5	46.6	4.9	.35
- feitelijke tijd rekenen	252.3	50.8	243.4	39.9	.56
. % van de totale instructietijd (C)	20.9	5.0	19.4	2.7	1.20
- feitelijke tijd lezen	137.4	49.0	126.2	42.1	.71
. % van de totale instructietijd (C)	11.2	3.7	10.0	3.3	1.03
- feitelijke tijd taal	182.6	32.6	215.0	43.8	-2.15 *
. % van de totale instructietijd (C)	15.1	2.3	17.1	3.5	-1.70
Jaargroep 8 :	(n=13)		(n=30)		
- feitelijke tijd rekenen-taal-lezen	554.1	67.9	557.1	116.4	- .11
. % van de totale instructietijd (C)	44.2	5.3	43.4	7.4	.26
- feitelijke tijd rekenen	247.1	42.7	217.2	50.8	1.85
. % van de totale instructietijd (C)	19.6	3.5	16.9	3.3	2.46 *
- feitelijke tijd lezen	105.1	38.1	119.4	47.0	- .97
. % van de totale instructietijd (C)	8.4	3.1	9.3	3.5	- .84
- feitelijke tijd taal	202.4	60.0	220.7	55.9	- .97
. % van de totale instructietijd (C)	16.0	4.5	17.2	3.8	- .90

C = instructietijd (gerealiseerde lestijd exclusief 'organisatie en 'overige activiteiten').
* $p < .05$ (tweezijdig).

Leerkrachtactiviteiten en groeperingsvormen tijdens lees-, taal- en rekenlessen
De derde centrale vraagstelling van deze studie betreft de verdeling van de beschikbare tijd binnen de lees-, taal- en rekenlessen over diverse instructie-activiteiten en groeperingsvormen. De beantwoording van deze vraag is gebaseerd op gegevens uit de lesoverzichten in het logboek. We gaan eerst in op de verdeling van de tijd over de verschillende instructie- of leerkrachtactiviteiten.

Leerkrachtactiviteiten. In de lesoverzichten (zie Figuur 1) werden de volgende categorieën leerkrachtactiviteiten onderscheiden: 'terugblik' (korte herhaling voorafgaande leerstof, huiswerkbespreking); 'instructie' (presentatie nieuwe leerstof); 'begeleid (in)oefenen' (gerichte controle op begrip nieuwe leerstof); 'individuele hulp' (de leerkracht geeft individuele begeleiding); 'nabespreking' (nabespreking van de nieuwe leerstof en opdrachten); 'herhaling' (herhaling van eerder geleerde leerstof); 'organisatie' (management-activiteiten en dergelijke) en 'pauze' (lesonderbreking). Voor combinatiegroepen was de categorie 'andere jaargroep' toegevoegd. Als de leerkracht tijdens de les met (leerlingen van de) niet bij het onderzoek betrokken jaargroep aan het werk was, hoefde niet ingevuld te worden wat hij/zij dan aan het doen was. Als de leerkracht binnen een les met de 'andere jaargroep' bezig is, werken de leerlingen uit de bedoelde groep zelfstandig, zonder hulp van de leerkracht.

In Tabel 3 is te zien hoe de lestijd verdeeld werd over de onderscheiden leerkrachtactiviteiten. Per leerkracht en per vak is uit de ingevulde lesoverzichten berekend hoeveel minuten elke activiteit gemiddeld over de (meestal 5) lessen voorkwam. Eventuele pauzes binnen de lessen zijn buiten beschouwing gelaten. In de tabel wordt de tijdsbesteding uitgedrukt in percentages van de totale lestijd, die steeds is omgerekend naar 40 minuten. De feitelijke gemiddelde lesduur voor rekenen, lezen en taal bleek niet te verschillen tussen combinatiegroepen en enkelvoudige groepen. De gegevens zijn voor de drie jaargroepen samen weergegeven om het aantal lessen voldoende groot te houden. Tussen de jaargroepen bleken overigens geen grote verschillen te bestaan (zie Nederlof, 1994).

Tijdens de *rekenlessen* werd gemiddeld het meeste tijd besteed aan het geven van individuele hulp, in combinatieklassen significant minder dan in enkelvoudige klassen (verschil gemiddeld 5 minuten per les). Het laatste gold ook voor begeleid (in)oefenen. Een groot gedeelte van dit verschil wordt waarschijnlijk verklaard door de tijd die leerkrachten in combinatieklassen besteedden aan de andere jaargroep. Gemiddeld richtten zij zich tijdens de rekenlessen 29% (12 min.) van de lestijd op de andere jaargroep. De leerlingen werken dan zonder hulp van de leerkracht. Gemiddeld 18% van de lestijd werd besteed aan het presenteren van nieuwe leerstof (instructie). Aan het geven van een terugblik of het herhalen van eerder geleerde leerstof werd het minste tijd besteed. Leerkrachten uit enkelvoudige groepen bleken significant meer tijd nodig te hebben (2% of 1 min.) voor niet-inhoudelijk gerichte leerkrachtactiviteiten (organisatie) dan leerkrachten in combinatieklassen.

Tijdens de *leeslessen* werd de meeste tijd besteed aan het begeleid (in)oefenen en het geven van individuele hulp, in enkelvoudige groepen iets meer dan in combinatiegroepen. Deze verschillen waren minder groot dan bij de rekenlessen en niet significant mede als gevolg van veel grotere standaardafwijkingen. In vergelijking met de rekenlessen werd tijdens de leeslessen gemiddeld duidelijk minder tijd besteed aan het presenteren van nieuwe leerstof (instructie), namelijk zo'n 6% minder (2 min.). Net als in de rekenlessen werd relatief weinig tijd besteed aan het geven van een terugblik of het herhalen van eerder geleerde leerstof. Gemiddeld rond 5% (2 min.) van de lestijd bleek nodig voor organisatie, in beide typen groepen.

Tijdens de *taallessen* werd gemiddeld ongeveer evenveel tijd aan 'instructie' besteed als in de rekenlessen (in beide gevallen ongeveer 18% of 7 min.). De tijd nodig voor organisatie lag, net als bij lezen, rond 5% (2 min.). Leerkrachten in enkelvoudige groepen besteedden significant meer tijd aan het begeleid (in)oefenen van nieuwe leerstof dan leerkrachten in combinatiegroepen (gemiddeld 13% of 5 min. meer).

De verschillen tussen leerkrachten in besteding van de lestijd waren groot. De standaardafwijkingen voor de tijd besteed aan begeleid (in)oefenen en individuele hulp lagen bij alle

Tabel 3. Verdeling van de lestijd over leerkracht-activiteiten per vak voor leerjaar 4, 6 en 8 samen (percentages).

	Enkelvoudige groepen gem.	sd.	*Combinatiegroepen* gem.	sd.	t	
Rekenen :	(n=35)		(n=83)			
- terugblik	4.8	3.8	3.7	4.1	1.44	
- instructie	16.9	7.8	18.6	7.8	-1.28	
- begeleid (in)oefenen	24.3	20.2	12.1	9.6	3.41	**
- individuele hulp	36.0	22.3	23.9	16.8	2.90	**
- nabespreking	7.9	7.4	6.1	6.4	1.35	
- herhaling	5.2	7.1	3.7	8.3	.93	
- andere jaargroep	-	-	29.1	18.1	-	
- organisatie	5.1	7.5	2.9	3.8	2.15	*
Lezen :	(n=35)		(n=85)			
- terugblik	4.2	9.3	3.6	5.8	.40	
- instructie	10.8	8.8	13.4	12.9	-1.07	
- begeleid (in)oefenen	38.5	27.4	30.3	25.5	1.58	
- individuele hulp	29.8	28.4	23.7	24.3	1.19	
- nabespreking	7.2	8.1	6.3	6.5	.62	
- herhaling	3.7	13.9	1.7	5.8	.83	
- andere jaargroep	-	-	16.5	20.0	-	
- organisatie	5.8	6.9	4.6	6.5	.94	
Taal :	(n=35)		(n=85)			
- terugblik	6.1	5.9	4.9	5.6	.99	
- instructie	18.1	10.2	18.8	9.8	- .34	
- begeleid (in)oefenen	28.2	20.9	15.0	12.3	3.48	**
- individuele hulp	25.7	22.0	20.0	16.3	1.38	
- nabespreking	9.7	7.3	7.4	8.1	1.45	
- herhaling	6.8	10.5	5.5	9.3	.67	
- andere jaargroep	-	-	24.0	19.1	-	
- organisatie	5.5	6.4	4.3	5.7	.93	

* $p < .05$; ** $p < .01$ (tweezijdig)

vakken tussen 20 en 29% in de enkelvoudige groepen, in combinatiegroepen bij lezen rond 25% en bij rekenen en taal tussen 10 en 16%. Ook de standaardafwijkingen met betrekking tot 'instructie' waren bij alle drie de vakken groot. Minimaal bedroeg deze standaardafwijking 8% (rekenen: enkelvoudige en gecombineerde groepen). De combinatiegroepen kwamen bij de leeslessen ten aanzien van deze categorie zelfs tot een standaardafwijking van 13%.

Groeperingsvorm. Naast de leerkrachtactiviteiten werden ook de gehanteerde groeperingsvormen geregistreerd. In de lesoverzichten is door de leerkrachten steeds aangegeven op welke doelgroep de activiteiten gericht waren. Daarbij werden voor de enkelvoudige groepen drie categorieën onderscheiden (klassikaal, subgroep en geen contact) en voor de combinatieklassen zes (klassikaal, hele jaargroep 4, 6 of 8, subgroep binnen, subgroep tussen, andere jaargroep en geen contact). 'Klassikaal' werd niet alleen gescoord als de leerkracht aan alle aanwezige leerlingen uitleg gaf, maar ook als de leerlingen individueel zelfstandig werkten en de leerkracht voor iedereen beschikbaar was voor hulp. In combinatiegroepen kreeg 'klassikaal' de betekenis: de leerlingen van alle aanwezige jaargroepen. In gecombineerde groepen wordt vaak afwisselend met leerlingen uit de ene en leerlingen uit de andere aanwezige jaargroep gewerkt. Daarom zijn de categorieën 'hele jaargroep (4, 6 of 8)' en 'andere jaargroep' opgenomen. Op

deze twee categorieën is dezelfde opmerking van toepassing als die ten aanzien van klassikaal werken is gemaakt. 'Subgroep' werd gescoord als de leerkracht gericht was op een groepje leerlingen uit de klas. Voor combinatiegroepen werd hierbij onderscheid gemaakt tussen een subgroep binnen een jaargroep ('subgroep binnen') en een jaargroepoverstijgende subgroep ('subgroep tussen'). 'Geen contact' ten slotte werd gescoord als de leerkrachten helemaal niet op de leerlingen gericht waren (ook niet beschikbaar voor hulp), maar met andere zaken bezig zoals het nakijken van huiswerk, een praatje maken met collega's en dergelijke.

Op basis van de lesoverzichten is de gemiddelde tijd die per les aan een bepaalde groeperingsvorm werd besteed, berekend. Deze gegevens, uitgedrukt in percentages van de totale lestijd (omgerekend naar 40 minuten), staan in Tabel 4, apart voor de reken-, lees- en taallessen.

Om de combinatiegroepen en enkelvoudige groepen op de categorie 'subgroep' te kunnen vergelijken, hebben we bij de combinatiegroepen de tijd die besteed werd aan 'subgroep binnen' en 'subgroep tussen' bij elkaar opgeteld. De hoeveelheid tijd die leerkrachten besteedden aan het werken met de hele klas vergelijken we op twee manieren. Eerst kijken we naar de tijd dat de leerkrachtactiviteiten op de hele klas gericht zijn (de categorie 'klassikaal'). Ten tweede nemen we voor combinatiegroepen de categorieën 'klassikaal' en 'hele jaargroep' samen.

Uit de tabel blijkt dat in enkelvoudige groepen bij alle vakken gemiddeld een groot deel van de tijd klassikaal werd gewerkt. Bij rekenen en taal lag het percentage klassikaal werken rond 85% (34 min.) en bij lezen rond 72% (29 min.). De overige tijd werd vooral besteed aan groepswerk (bij rekenen en taal tussen 11 en 14%, ofwel 4 à 5 minuten en bij lezen 23% ofwel ruim 9 minuten). In enkelvoudige klassen werd dus weinig in subgroepen gewerkt, hoewel bij lezen meer dan bij de andere vakken (niveau-lezen). Het percentage tijd dat leerkrachten uit de

Tabel 4. Verdeling van de lestijd naar doelgroepcategorieën per vak voor de leerjaren 4, 6 en 8 samen (percentages).

	Enkelvoudige groepen		*Combinatiegroepen*		
	gem.	sd.	gem	sd.	t
Rekenen :	(n=35)		(n=83)		
klassikaal	83.8	22.0	16.6	20.4	15.98 **
hele jaargroep	-	-	46.3	26.0	-
klassikaal + hele jaargroep	83.8	22.0	62.9	21.0	4.87 **
subgroep	13.8	22.0	8.4	14.8	1.32
geen contact	2.4	5.4	1.1	2.6	1.42
andere jaargroep	-	-	27.6	17.9	-
Lezen :	(n=35)		(n=85)		
klassikaal	72.3	25.4	36.3	33.1	6.43 **
hele jaargroep	-	-	31.2	25.3	-
klassikaal + hele jaargroep	72.3	25.4	67.5	25.4	.94
subgroep	23.1	23.2	15.0	21.1	1.86
geen contact	4.6	8.9	1.0	3.8	2.30 *
andere jaargroep	-	-	16.5	20.0	-
Taal :	(n=35)		(n=85)		
klassikaal	85.4	18.4	25.6	27.4	14.12 **
hele jaargroep	-	-	43.7	24.1	-
klassikaal + hele jaargroep	85.4	18.4	69.3	21.1	3.94 **
subgroep	10.7	19.0	5.6	12.8	1.45
geen contact	3.9	7.0	1.3	4.2	2.09 *
andere jaargroep	-	-	23.8	18.7	-

* $p < .05$; ** $p < .01$ (tweezijdig)

enkelvoudige groepen in het geheel niet op de leerlingen gericht waren ('geen contact'), lag bij alle vakken gemiddeld vrij laag, namelijk tussen de 2 en 5% (1 à 2 minuten).

In combinatiegroepen werd geheel overeenkomstig de verwachtingen bij alle vakken veel minder klassikaal gewerkt dan in enkelvoudige groepen (steeds significant). Tijdens de rekenlessen werd het minst klassikaal gewerkt (17% of 7 min.), tijdens de taallessen wat meer (26% of 10 min.) en tijdens de leeslessen het meest (36% of 15 min.). Waarschijnlijk lenen leeslessen zich wat beter voor jaargroepoverstijgend werken dan taal- en vooral rekenlessen. Bij alle vakken werd in combinatiegroepen minder (niet significant) in subgroepen gewerkt dan in enkelvoudige groepen. Als we kijken naar de tijd die in combinatiegroepen met de hele bedoelde jaargroep werd gewerkt, was dat bij rekenen en taal gemiddeld bijna de helft van de les en bij lezen bijna een derde van de les. Wanneer we voor de combinatiegroepen de tijd die werd besteed aan klassikaal werken en werken met de hele jaargroep bij elkaar nemen, dan is het verschil met de tijd die gemiddeld in enkelvoudige groepen klassikaal werd gewerkt bij rekenen en taal nog steeds significant, bij lezen niet meer. Bij rekenen en taal werd rond een kwart van de lestijd in combinatiegroepen aan de andere jaargroep (niet 4, 6 of 8) besteed en bij lezen minder, namelijk 17% van de lestijd. Dit hangt samen met het eerder vermelde relatief grotere aandeel van klassikaal werken bij lezen.

Leerkrachten in combinatiegroepen waren gemiddeld bij alle vakken slechts 1% van de lestijd niet op de leerlingen gericht. Dit is significant minder dan in enkelvoudige groepen, althans bij lezen en taal (enkelvoudige groepen resp. 5% en 4% 'geen contact'). Leerkrachten in enkelvoudige groepen lijken tijdens de lessen wat makkelijker tijd te vinden voor het even maken van een praatje met een collega, vast wat spullen voor de volgende les klaarzetten, etcetera.

CONCLUSIES EN DISCUSSIE

In deze studie stond de vraag centraal op welke wijze de beschikbare onderwijstijd in combinatieklassen en enkelvoudige klassen wordt besteed. Drie aspecten hiervan hebben we onderzocht, namelijk (1) de verhouding tussen bruto onderwijstijd (roostertijd), feitelijk gerealiseerde tijd en instructietijd (gerealiseerde onderwijstijd exclusief organisatie en overige activiteiten), (2) de beschikbare instructietijd voor de vakken rekenen, lezen en taal en ten slotte (3) de verdeling van de tijd binnen reken-, lees- en taallessen over instructie-activiteiten en groeperingsvormen.

Enkelvoudige groepen en combinatiegroepen hebben gemiddeld per week evenveel roostertijd ter beschikking, 22 uur in groep 4 en 24½ uur in groep 6 en 8. De verschillen tussen klassen zijn klein; voor groep 4 is de standaardafwijking ongeveer twee keer zo groot als voor de groepen 6 en 8. Zowel enkelvoudige als gecombineerde groepen blijken vrijwel alle roostertijd ook feitelijk te realiseren. Het gemiddelde tijdverlies per dag blijft beperkt tot 5 à 6 minuten. Het verlies blijkt met name veroorzaakt door enkele minuten later beginnen, iets eerder opruimen, iets langer pauze nemen, en dergelijke. Deze gegevens komen overeen met de gegevens in het CEB-rapport (1994), hoewel opgemerkt moet worden dat onze cijfers wat geflatteerd kunnen zijn vanwege de gegevensverzameling in zo 'normaal' mogelijke weken en vanwege de relatief korte periode waarin het logboek is bijgehouden (twee weken). Het tijdschrijven kan immers tot gevolg hebben dat men zich beter aan de roostertijden houdt dan normaal en de kans hierop is groter naarmate de logboekperiode korter is.

In het CEB-rapport worden ook verschillen tussen leerkrachten wat betreft de realisatie van de roostertijd vermeld. Bij 81% van de leraren blijft het tijdverlies beperkt tot een kwartier per dag. Zo'n 9% van de leerkrachten houdt zich exact aan de lestijden. Ongeveer 11% weet de leertijd met 5 minuten per dag te verhogen. In incidentele gevallen loopt de verlenging van de schooltijd op tot meer dan een half uur (o.c., p. 37-38). Wij vonden vergelijkbare verschillen tussen leerkrachten onderling. Voor 75% van de leerkrachten ligt het percentage gerealiseerde lestijd boven 96.5%, wat neerkomt op een maximaal tijdverlies per dag van 10 minuten.

De gerealiseerde onderwijstijd met aftrek van tijd nodig voor 'organisatie' en 'overige acti-

viteiten' noemden we de instructietijd. Deze bedraagt gemiddeld 85% van de gerealiseerde lestijd. Gemiddeld wordt ongeveer 10% van de gerealiseerde lestijd aan organisatie besteed. Hierbij moeten we opmerken dat leerkrachten de tijd voor 'organisatie' en 'overige activiteiten' niet hoefden te registreren, wanneer deze minder dan één minuut in beslag nam. Daarnaast zou het zo kunnen zijn dat leerkrachten juist door het invullen van het logboek gestimuleerd werden tot efficiënter organiseren. Het beeld zou in werkelijkheid dus nog wat ongunstiger kunnen zijn.

Een opvallende bevinding is dat leerkrachten in combinatiegroepen gemiddeld minder tijd besteden aan organisatie dan leerkrachten uit enkelvoudige groepen. In jaargroep 6 blijkt dit verschil statistisch significant. De verwachting lijkt ons gerechtvaardigd dat de klasse-organisatie van leerkrachten in combinatiegroepen onder grotere druk staat dan die van leerkrachten uit enkelvoudige groepen. Van leerkrachten in de combinatieklas wordt verwacht dat zij in staat zijn toezicht te houden op beide jaargroepen. De leerkrachten moeten hun aandacht goed over verschillende zaken tegelijk kunnen verdelen. Mogelijk bereiden leerkrachten van combinatiegroepen hun lessen beter voor, daartoe gedwongen door de complexere organisatie van het onderwijs in hun groep. Betere voorbereiding kan leiden tot efficiënter organiseren tijdens de lessen. Deze verklaring spoort met de bevinding uit een enquête van de Inspectie Basisonderwijs uit 1978 (zie Roelofs, 1993), waarin leerkrachten onder andere aangaven dat de combinatieklas meer tijd aan voorbereiding vraagt. De combinatieklas eist een grotere inzet van de leerkracht. Dat het efficiënt organiseren van het onderwijs hen niet altijd makkelijk valt blijkt uit interviewgegevens van Veenman et al. (1986). Heel sterk wordt de zware taakbelasting van leerkrachten in de combinatieklas benadrukt. Uit de interviewgegevens blijkt dat het de leerkracht erg veel tijd en energie kost om een combinatieklas goed draaiende te houden. Een zware wissel wordt getrokken op zijn of haar leiderschaps- en managementkwaliteiten. Uit kwalitatieve observaties van Veenman et al. (1986) komt naar voren dat een aantal managementvaardigheden die in de literatuur als effectief worden onderscheiden vooral voor het omgaan met combinatieklassen van belang zijn.

Er zijn nauwelijks verschillen in gemiddelde hoeveelheid instructietijd tussen enkelvoudige en gecombineerde groepen; de verschillen tussen leerkrachten (scholen) blijken echter erg groot te zijn. De standaardafwijking voor verschillen in instructietijd is twee tot vier keer zo groot als die voor de gerealiseerde lestijd. Sommige leerkrachten hebben meer tijd voor organisatie nodig dan andere, als gevolg van verschillen in beschikbare voorzieningen en als gevolg van meer of minder efficiënt klasse-management.

Samenvattend kunnen we concluderen dat combinatiegroepen (in kleine scholen) gemiddeld ten minste evenveel of iets meer instructietijd realiseren dan enkelvoudige groepen (in grote scholen). Zeker ten aanzien van de leerkrachten in combinatiegroepen moeten we dit positief waarderen. Verder kunnen we concluderen dat er belangrijke verschillen bestaan tussen leerkrachten (scholen) ten aanzien van efficiënt gebruik van de beschikbare tijd.

De *gemiddelde tijd besteed aan de vakken rekenen, lezen en taal* verschilt weinig tussen enkelvoudige en gecombineerde groepen, maar er blijken wel grote verschillen tussen klassen te zijn. Gemiddeld wordt ongeveer 19% van de instructietijd aan rekenen besteed en ongeveer 27% aan taal en lezen. In literatuur over effectieve scholen wordt een accent op de verwerving van basisvaardigheden veelal als één van de effectiviteitsfactoren genoemd (zie bijvoorbeeld Levine & Lezotte, 1990). Uit ons onderzoek blijkt dat zowel enkelvoudige als gecombineerde groepen gemiddeld ongeveer 46.5% van de instructietijd aan de vakken rekenen, lezen en taal besteden. Of dit voldoende tijd is om de kerndoelen ten aanzien van deze basisvaardigheden te realiseren is onduidelijk. Hierover is momenteel onvoldoende informatie beschikbaar (CEB, 1994).

Concluderend kunnen we ten aanzien van de tweede onderzoeksvraag stellen dat we geen belangrijke verschillen tussen enkelvoudige groepen en combinatiegroepen kunnen waarnemen met betrekking tot de tijdsbesteding aan de vakken rekenen, lezen en taal. Wel zien we vrij grote verschillen tussen leerkrachten (scholen) onderling. Als we de standaardafwijkingen relatief ten opzichte van de gemiddelden uitdrukken, dan kunnen we constateren dat de verschillen tussen leerkrachten (scholen) toenemen, naarmate we meer specifiek naar invulling van de beschikbare

tijd gaan kijken: roostertijd – gerealiseerde totale lestijd – totaal beschikbare instructietijd – instructietijd beschikbaar voor de vakken rekenen, lezen en taal.

Wat betreft de *instructie-activiteiten* zien we dat de leerkrachten in enkelvoudige groepen bij alle drie de vakken gemiddeld meer tijd besteden aan het begeleid (in)oefenen van nieuwe stof en aan het geven van individuele hulp dan leerkrachten in combinatiegroepen. De waargenomen verschillen zijn steeds groter dan 5%, maar niet voor alle drie de vakken statistisch significant. De verschillen zijn te verklaren uit het feit dat de leerkracht in combinatiegroepen zijn (instructie)tijd moet verdelen over twee of meer jaargroepen. Gemiddeld werken leerlingen in combinatiegroepen meer tijd zonder de hulp van de leerkracht. In combinatiegroepen lijkt hiermee een grotere nadruk op zelfstandige verwerking van de leerstof te liggen. De leerling in de combinatiegroep lijkt zich daardoor in een wat meer nadelige positie te bevinden. Veenman et al. (1986) hebben in hun observatiestudie eveneens gegevens verzameld over de hoeveelheid tijd die de leerkracht besteedt aan verschillende leerkrachtactiviteiten tijdens de lessen, rekenen, lezen en taal. In hoofdlijnen komen de resultaten overeen. Veenman et al. kwamen tot de conclusie dat leerlingen in enkelvoudige klassen meer individuele hulp, supervisie en groepsgewijze instructie kregen, vergeleken met leerlingen in combinatiegroepen.

Ten aanzien van de *managementactiviteiten* (organisatie) tijdens de les leiden de logboekgegevens echter tot andere bevindingen dan de observaties van Veenman et al. (1986). Uit onze gegevens blijken leerkrachten gemiddeld weinig tijd (ongeveer 5%) te 'verliezen' aan organisatie binnen de lessen. Leerkrachten in combinatiegroepen blijken gemiddeld iets minder tijd aan organisatie kwijt te zijn dan leerkrachten in enkelvoudige groepen. Uit het onderzoek van Veenman et al. blijkt dat leerkrachten gemiddeld 10% van de lestijd aan managementactiviteiten besteden. Verder kwamen zij tot de conclusie dat de leerkracht uit de combinatieklas gemiddeld iets meer tijd besteedt aan managementactiviteiten tijdens de lessen rekenen, lezen en taal, dan de leerkracht uit de enkelvoudige klas. In beide onderzoeken zijn de verschillen tussen de twee typen klassen klein. We kunnen dan ook veilig concluderen dat er geen belangrijke verschillen in organisatietijd binnen de lessen rekenen, lezen en taal bestaan tussen beide soorten groepen. Het verschil tussen beide studies in gemiddelde omvang van de categorie 'organisatie' kan een gevolg zijn van het verschil in onderzoeksmethode.

Wat betreft de *groeperingsvormen* blijken leerkrachten in enkelvoudige groepen gemiddeld significant meer tijd te besteden aan klassikaal werken dan leerkrachten in combinatiegroepen. Dit is uiteraard een direct gevolg van de aanwezigheid van twee of meer jaargroepen in een combinatieklas. Wanneer we voor de combinatiegroepen de categorieën 'klassikaal' en 'hele jaargroep' samen nemen dan blijkt het verschil in tijdsbesteding tussen enkelvoudige en gecombineerde groepen bij zowel rekenen als taal nog steeds significant. Er wordt nauwelijks in subgroepen gewerkt, noch in de enkelvoudige groepen, noch in de combinatiegroepen. In combinatiegroepen wordt relatief veel tijd gewerkt met de hele jaargroep. Tot slot blijkt uit onze onderzoeksgegevens dat leerkrachten in enkelvoudige groepen gemiddeld iets meer tijd (1 à 3%) in het geheel niet op de leerlingen gericht zijn dan leerkrachten uit combinatiegroepen. Voor lees- en taallessen is dit verschil significant.

Deze gegevens verkregen met de logboeken blijken in grote lijnen weer overeen te komen met conclusies uit het observatieonderzoek van Veenman et al. (1986). Zij kwamen tot de conclusie dat leerlingen in enkelvoudige klassen meer klassikaal onderwezen worden dan leerlingen in combinatieklassen. Ten aanzien van het werken in subgroepen werd door Veenman et al. geconcludeerd dat het werken in subgroepen aan een gemeenschappelijke taak geen gangbare praktijk is in het onderwijs. Volgens deze auteurs zitten de leerlingen wel in subgroepen, maar binnen deze subgroepen is iedere leerling voor zich bezig. Verder blijkt uit het genoemde onderzoek dat de leerkracht in de combinatieklas minder rustmomenten en minder gelegenheid om eigen activiteiten te verrichten heeft, bijvoorbeeld voor de voorbereiding op de volgende les, dan de leerkracht in de enkelvoudige klas.

Zowel uit de resultaten van Veenman et al. (1986), als uit onze gegevens kan afgeleid worden dat de jaarklassikale situatie in de combinatiegroep wordt voortgezet. De jaargroepen functio-

neren als het ware als gescheiden groepen, die om beurten les ontvangen. In dit opzicht blijkt de constatering van de Inspectie Basisonderwijs uit 1978 15 jaar na dato nog steeds van kracht: 'De mogelijkheden van bijvoorbeeld het werken in niveaugroepen en andere vormen van differentiatie en individualisering en een meer flexibeler schoolorganisatie worden zeker te weinig onderkend. Te veel blijft men werken in de jaarklassikale situatie' (geciteerd in Roelofs, 1993). Ter volledigheid moet uit onze onderzoeksgegevens worden toegevoegd dat tijdens de leeslessen de jaarklassikale situatie enigszins wordt doorbroken. Waarschijnlijk hangt dit samen met het zogenaamde niveaugroepen-lezen (AVI-lezen).

In tijd die op school beschikbaar is om te leren en in verdeling van de lestijd over instructieactiviteiten en groeperingsvormen bestaan belangrijke en soms aanzienlijke verschillen tussen klassen (leerkrachten). Er zijn op basis van het kwantitatieve aanbod echter geen redenen om aan te nemen dat combinatieklassen minder effectief zullen zijn dan enkelvoudige klassen, of dat kleine scholen minder effectief zullen zijn dan grote scholen.

LITERATUUR

Anderson, L.W. (1994). Time, allocated and instructional. In T. Husén, & T.N. Postlethwaite (eds.). *The International encyclopedia of education* (2nd edition) (pp. 6388-6390). Oxford: Pergamon.

Arlin, M. (1979). Teacher transitions can disrupt time flow in classrooms. *American Educational Research Journal, 16*, 42-56.

Commissie Evaluatie Basisonderwijs (1994). *Inhoud en opbrengsten van het basisonderwijs*. De Meern: Inspectie van het Onderwijs.

Elley, W.B. (Ed.) (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford: Pergamon.

Fisher, C., Filby, N., Marliave, R., Cahen, L., Dishaw, M., Moore, J., & Berliner, D. (1978). *Teaching behaviors, academic learning time, and student achievement.* (Beginning teacher evaluation study. Phase III-B, final report). San Francisco: Far-West Laboratory.

Fredrick, W.C. & Walberg, H.J. (1980). Learning as a function of time. *Journal of Educational Research, 73*, 183-194.

Gettinger, M. (1989). Effects of maximizing time spent and minimizing time needed for learning on pupil achievement. *American Educational Research Journal, 26*, 73-91.

Harskamp, E. & Suhre, C. (1992). Psychometrische kwaliteiten van meetinstrumenten van het onderwijsaanbod. *Tijdschrift voor Onderwijsresearch, 17*, 355-369.

Harskamp, E.G., Suhre, C.J.M., & Willemsen, T.F.W.P. (1991). *De kwaliteit van meetinstrumenten voor het aanbod van rekenen-wiskunde*. Groningen: RION.

Heidelberg, J., Bergh, H. van den, Kuhlemeier, H., & Hoeks, J. (1993). Tijdsbesteding van docenten in het voortgezet onderwijs op verschillende manieren gemeten: een validatiestudie. *Tijdschrift voor Onderwijsresearch, 18*, 306-316.

Kral, M. (in voorbereiding). *Effecten van schoolgrootte. Een onderzoek naar effectiviteitsverschillen tussen combinatieklassen en enkelvoudige klassen in het basisonderwijs* (voorlopige titel). Nijmegen: Instituut voor Toegepaste Sociale wetenschappen.

Leinhardt, G. (1987). Types of content. In M.J. Dunkin (ed.). *The international encyclopedia of teaching and teacher education* (pp. 359-364). Oxford: Pergamon.

Lem, P., Veenman, S., Nijssen, F., & Roelofs, E. (1988). *Omgaan met combinatieklassen. Verslag van een teamgerichte nascholing*. 's Gravenhage: Instituut voor Onderzoek van het Onderwijs.

Levine, D.U. & Lezotte, L.W. (1990). *Unusually effective schools: a review and analyses of research and practice*. Madison: National Center for Effective Schools.

Nederlof, W. (1994). *Tijd en leren*. Doctoraalscriptie Onderwijskunde. Open Universiteit, Heerlen.

Otter, M.E. (1993). *Leesvaardigheid, leesonderwijs en buitenschools lezen.* (Proefschrift Universiteit van Amsterdam). Amsterdam: SCO.

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*, 281-287.

Overmaat, A.M. (1992). *De kwaliteit van meetinstrumenten voor het aanbod van taal en lezen*. Amsterdam: SCO.

Roelofs, E. (1993). *Teamgerichte nascholing en coaching*. Nijmegen: Proefschrift K.U. Nijmegen.

Rosenshine, B. & Stevens, R. (1986). Teaching functions. In M.C. Wittrock (ed.). *Handbook of research on teaching* (3rd ed., pp. 376-391). New York: MacMillan.

Veenman, S., P. Lem, E. Roelofs, & F. Nijssen (1992). *Effectieve instructie en doelmatig klassemanagement: een schoolverbeteringsprogramma voor enkelvoudige en combinatieklassen.* Amsterdam/ Lisse: Swets en Zeitlinger.

Veenman, S., Lem, P., Voeten, M., Winkelmolen, G., & Lassche H. (1986). *Onderwijs in combinatieklassen.* 's Gravenhage: Instituut voor Onderzoek van het Onderwijs.

Veenman, S. (1987). Leertijd en basisvorming. In J. van Kuyk (red.). *Basisvorming in de basisschool* (pp. 246-264). Tilburg: Zwijsen.

Walberg, H.J. (1986). Synthesis of research on teaching. In M.C. Wittrock (ed.). *Handbook of Research on Teaching* (3rd ed., pp. 214-229). New York: Macmillan.

Wijnstra, J.M. (1988). *Balans van het rekenonderwijs in de basisschool.* Uitkomsten van de eerste rekenpeiling medio en einde basisonderwijs. Arnhem: CITO.

Wisse, M. (1993). *Verslag van de totstandkoming van het logboek bij 'Onderwijs, Instructie en Leerresultaten in combinatiegroepen'.* Doctoraalscriptie Onderwijskunde. Open Universiteit, Heerlen.

Zwarts, M. (Red.) (1990). *Balans van het taalonderwijs aan het einde van de basisschool* (PPON-reeks nr. 2). Arnhem: Cito.

Manuscript ontvangen 7-11-1994
Definitieve versie ontvangen 6-4-1995

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 4, pp. 318-331

Is strategische hulp effectief in een aardrijkskunde-computerles?

Teatske van der Zijpp
Centrum voor Educatieve Geografie

ABSTRACT

In this article the effects of two computermodules on mapskills of 113 seventh-graders in secondary schools were compared. It was investigated whether metacognitive help in computer modules concerning mapskills (= strategic help), resulted in more positive learning effects than a computer module without this strategic help. In this study students in the group with strategic help performed significantly better on the post-test then students in the group without strategic help. However, differences in learning effects of the two modules were modest. No interaction was found between ability of students on the pre-test of map skills and condition (module with or without strategic help). Finally, although, some significant differences were found between boys and girls on the pre-test, on the post-test no gender differences could be detected.

INLEIDING

Met de invoering van de basisvorming zijn er tal van veranderingen in het onderwijs doorgevoerd. Niet alleen in het onderwijsaanbod (met nieuwe vakken zoals verzorging en techniek) maar ook in de onderwijsaanpak. Een sterk accent wordt in de kerndoelen van de basisvorming gelegd op vaardigheden (Handboek Basisvorming, III 2c.voo.1). Ook voor de tweede fase van het voortgezet onderwijs is er een rapport verschenen waarin staat dat meer nadruk op vaardigheden is geboden (Stuurgroep Profiel Tweede Fase Voortgezet Onderwijs, 1994). De term vaardigheden (of: procedurele kennis) wordt opgevat als het vermogen om iets te *kunnen* doen met de kennis waarover men beschikt (Anderson, 1990, p.219, Boekaerts en Simons, 1993, p.8). De Commissie Herziening Eindtermen heeft naast een zestal algemene vaardigheden (waaronder o.a. het verrichten van een eenvoudig onderzoek, het verwoorden van een eigen standpunt, het interactief leren samenwerken aan opdrachten) op het niveau van de afzonderlijke vakken, vakspecifieke vaardigheden geformuleerd. Onderwerp van dit artikel zijn vaardigheden die in de kerndoelen voor het vak aardrijkskunde als kaartvaardigheden staan omschreven. De centrale probleemstelling van dit onderzoek is: welke vakinhoudelijke eisen moeten worden gesteld aan lesmateriaal dat beoogt de kaartvaardigheden van zoveel mogelijk leerlingen in het voortgezet onderwijs te bevorderen?

Algemene, domeinspecifieke en metacognitieve vaardigheden

Voordat het begrip kaartvaardigheid wordt toegelicht, zal worden ingegaan op onderwijstheorieën ten aanzien van het leren van vaardigheden.

Naar de wijze waarop vaardigheden het beste aangeleerd kunnen worden zijn tal van studies gedaan. In het WRR-rapport "Jeugd in ontwikkeling" worden drie stromingen van studies naar leren denken en leren leren onderscheiden: de generalistische, de domein-specifieke en de gecombineerde benadering. "In de generalistische opvatting bestaan er algemene probleemoplossingsvaardigheden, denkvaardigheden en leervaardigheden die in allerlei verschillende situaties kunnen worden ingezet" (WRR-rapport, 1992, p.212). Uit tal van onderzoeken is echter gebleken dat transfer van vaardigheden uit het ene domein naar het andere weinig optreedt (zie

Adres: Centrum voor Educatieve Geografie IDO/VU, Vrije Universiteit Amsterdam, De Boelelaan 1115, 1081 HV Amsterdam.

Simons & Verschaffel, 1992). Bovendien wijst onderzoek er op dat een bepaalde hoeveelheid domeinkennis een voorwaarde is om efficiënt en effectief binnen een domein te kunnen presteren (zie o.a. Chase & Simon, 1973; Goldstein & Papert, 1977) "Goede probleemoplossers weten in vergelijking met slechte probleemoplossers meer over het vakgebied; kunnen vakspecifieke probleemtypen herleiden tot andere meer simpele problemen; hebben een categoriseringssysteem voor typen van problemen; en hebben een betere en een snellere toegang tot de relevante, benodigde kennis in hun geheugen" (WRR-rapport: Jeugd in ontwikkeling, 1992, p. 214). De domeinspecifieke benadering legt daarom de nadruk op hetzij het opbouwen van een kennisbestand en/of het aanleren van domeinspecifieke procedures. Ondanks onderzoeksresultaten die de domeinspecifieke benadering steunden, is men blijven zoeken naar mogelijkheden om algemene vaardigheden aan te leren die binnen verschillende domeinen kunnen worden gebruikt. Twee citaten illustreren de vasthoudendheid van onderzoekers aan de mogelijkheid van het succesvol aanleren van algemene vaardigheden aan leerlingen. "Of course, to appreciate the dependence of general skills application on specific knowledge is not to deny that such general skills exist. Yet such an understanding raises questions about the wisdom of attempting to develop thinking skills outside the context of specific knowledge domains. It suggests that a more promising route may be to teach thinking skills within specific disciplines and perhaps hope for some transfer to other disciplines as relevant knowldege is acquired" (Resnick, 1987, p.18). Perkins en Salomon formuleren het als volgt: "There are general cognitive skills; but they always function in contextualized ways" (Perkins & Salomon, 1989, p.19). Perkins en Salomon halen enkele studies aan van Schoenfeld (1985) en Palincsar & Brown (1984) waaruit is gebleken dat instructie in algemene vaardigheden zeer succesvol kan zijn mits in een domeinspecifieke situatie aangeboden en voorzien van vragen die zogenaamde metacognitieve vaardigheden uitlokken (Perkins & Salomon, 1989, p.20). Met metacognitieve vaardigheden of -strategieën (Flavell, 1979; Perkins & Salomon, 1989; Boekaerts & Simons, 1993), regulatieve vaardigheden (Vermunt, 1992) of strategische vaardigheden (Van Westrhenen, 1992), wordt gedoeld op het zelf actief sturing geven aan de eigen cognitieve processen door middel van oriënteren, plannen, bewaken, toetsen, herstellen, evalueren en reflecteren (Boekaerts & Simons, 1993, p.85). Deze vaardigheden zijn in tegenstelling tot domeinspecifieke vaardigheden niet betrokken op een bepaalde leerinhoud maar op de wijze waarop een leerling bij het uitvoeren van vakspecifieke taken zijn leren inricht en reguleert (Van Westrhenen, 1992, p.8). Een voorbeeld van een metacognitieve vaardigheid is het hardop omschrijven wat iemand zojuist heeft gezegd om te controleren of dat hetgene is dat deze persoon bedoelde (Flavell, 1979, p.909).

Gezien de succesvolle resultaten van de trainingsstudies van Palincsar & Brown (1984), Scardamalia, Bereiter & Steinbach (1984) en Schoenfeld (1985) met metacognitief trainingsmateriaal in een domeinspecifieke context die zich o.a. in spectaculaire prestatieverbeteringen in het domein zelf te zien gaven, wordt in het WRR-rapport een 'gecombineerde benadering' aanbevolen . De 'gecombineerde benadering' gaat er van uit dat een vakspecifieke kennisbasis essentieel is, maar dat algemene heuristieken en probleemoplossings-strategieën wel degelijk in de beschouwing moeten worden betrokken en dat met name de metacognitieve kennis en vaardigheden een bepalende rol kunnen spelen (WRR-rapport, 1992, p.214).

Omdat uit literatuur blijkt dat metacognitieve vaardigheden goede voorspellers zijn van leerprestaties (Wang, Haertel & Walberg, 1990, p.36; De Jong, 1992, p.180) en daarnaast aangetoond is dat 'metacognitieve vaardigheden niet uitsluitend onderdeel van intelligentie uitmaken' (Veenman & Elshout, 1992, p.299), lijkt het zinvol om in lesmateriaal dat gericht is op het leren van kaartvaardigheden ook aandacht te besteden aan deze metacognitieve vaardigheden. Kenmerkend voor de trainingsprogramma's van Palincsar & Brown (1984), Scardamalia, Bereiter & Steinbach (1984) en Schoenfeld (1985) is echter dat ze een vrij (arbeids)intensieve rol van de docent vereisen. In de dagelijkse onderwijspraktijk kan dit een probleem zijn. Een alternatief biedt het gebruik van computers als instructiemedium. Volgens Collins & Brown (1988) is de computer een heel geschikt middel om metacognitieve vaardigheden bij leerlingen te stimuleren. "Our claim here is that the computational medium, properly structured, can

provide a powerful, motivating, and as yet untapped tool for focusing the students' attention directly on their own thought processes" (Collins & Brown, 1988, p. 2). Succes is echter niet altijd verzekerd als de computer wordt gebruikt als instructiemedium voor het stimuleren van metacognitief handelen (zie bijvoorbeeld het gebrek aan resultaten met metacognitief computer-trainingsmateriaal onder ilo- en lto-leerlingen in een onderzoek door De Jong (1992). In het hieronder beschreven experiment is een poging gedaan om metacognitief computer-lesmateriaal te ontwerpen en te testen op de effectiviteit ervan op leerprestaties in het domein kaartvaardigheden binnen het aardrijkskunde-onderwijs.

Algemene en metacognitieve vaardigheden in het aardrijkskunde-onderwijs

Om te bepalen hoe metacognitieve handelingen in een aardrijkskunde-computerles bij het leren van kaartvaardigheden kunnen worden ingebracht zal eerst het domeinspecifieke element van aardrijkskunde moeten worden toegelicht. Hoekveld-Meijer & Schutte (1988, p.8) definiëren geografische kennis als kennis van regio's (gebieden) in hun onderlinge samenhang. Hiermee wordt niet alleen de kennis over wat er op een bepaalde plaats of in een gebied te vinden is bedoeld maar ook de kennis over hoe een plaats of gebied is gelegen ten opzichte van andere plaatsen of gebieden (Van der Schee, 1987, p.11). Kenmerkend voor de aardrijkskunde is dus het object van studie: namelijk gebieden, dit in tegenstelling tot andere maatschappijvakken zoals sociologie of psychologie die de mens als object van studie hebben.

Geografische vaardigheden zijn vaardigheden om verschijnselen dusdanig te ordenen dat deze als kenmerken van gebieden en relaties tussen gebieden kunnen worden onderscheiden. Kaarten zijn hierbij een onontbeerlijk geografisch hulpmiddel. "Een kaart plaatst ruimtelijke gegevens, dat wil zeggen gegevens over objecten of verschijnselen waarvan men de plaats op aarde kent, in hun onderlinge ruimtelijke relatie. Een kaart kan men beschouwen als een ruimtelijke informatiesysteem dat op vele vragen over het afgebeelde gebied antwoord geeft: de afstanden tussen punten ten opzichte van elkaar, de grootte van oppervlakken en de aard van spreidingspatronen. De antwoorden zijn meestal direct af te lezen uit het kaartbeeld, zonder dat er eerst iets moet worden ingetoetst of opgezocht (Ormeling, 1990, p.1)". Omdat kaartvaardigheden in de geografie zo'n belangrijk gereedschap bieden om geografische kennis (inzicht in de ruimtelijke samenhang van verschijnselen) te verwerven staan deze vaardigheden in de kerndoelen expliciet vermeld bij de vakspecifieke vaardigheden (Handboek Basisvorming, III.II.KER.I).

In literatuur over kaartgebruik (Muercke, 1978; Olson, 1979 en Van der Schee, 1987), wordt doorgaans onderscheid gemaakt tussen drie typen kaartvaardigheden: kaartlezen, kaartanalyse en kaartinterpretatie (zie schema 1). Zoals schema 1 laat zien, kunnen kaartvaardigheden, overeenkomstig de opvatting van Perkins en Salomon (1989, p.19), worden opgevat als algemene vaardigheden die in een specifiek domein (aardrijkskunde) en met een vakspecifiek leermiddel (kaarten) worden toegepast. De cognitieve vaardigheden (identificeren, classificeren en relateren) worden ook toegepast in de klassieke conceptleertaak van Bruner et. al. (1988). In deze conceptleertaak moeten de proefpersonen door middel van identificeren en classificeren (vergelijken) een regel ontdekken uit een reeks geometrische patronen die variëren in bijvoorbeeld vorm, grootte of kleur.

Van der Schee omschrijft kaartlezen als "het identificeren en benoemen van verschijnselen op een kaart". Kaartanalyse betreft "het onderkennen van ruimtelijke spreidingen, ruimtelijke

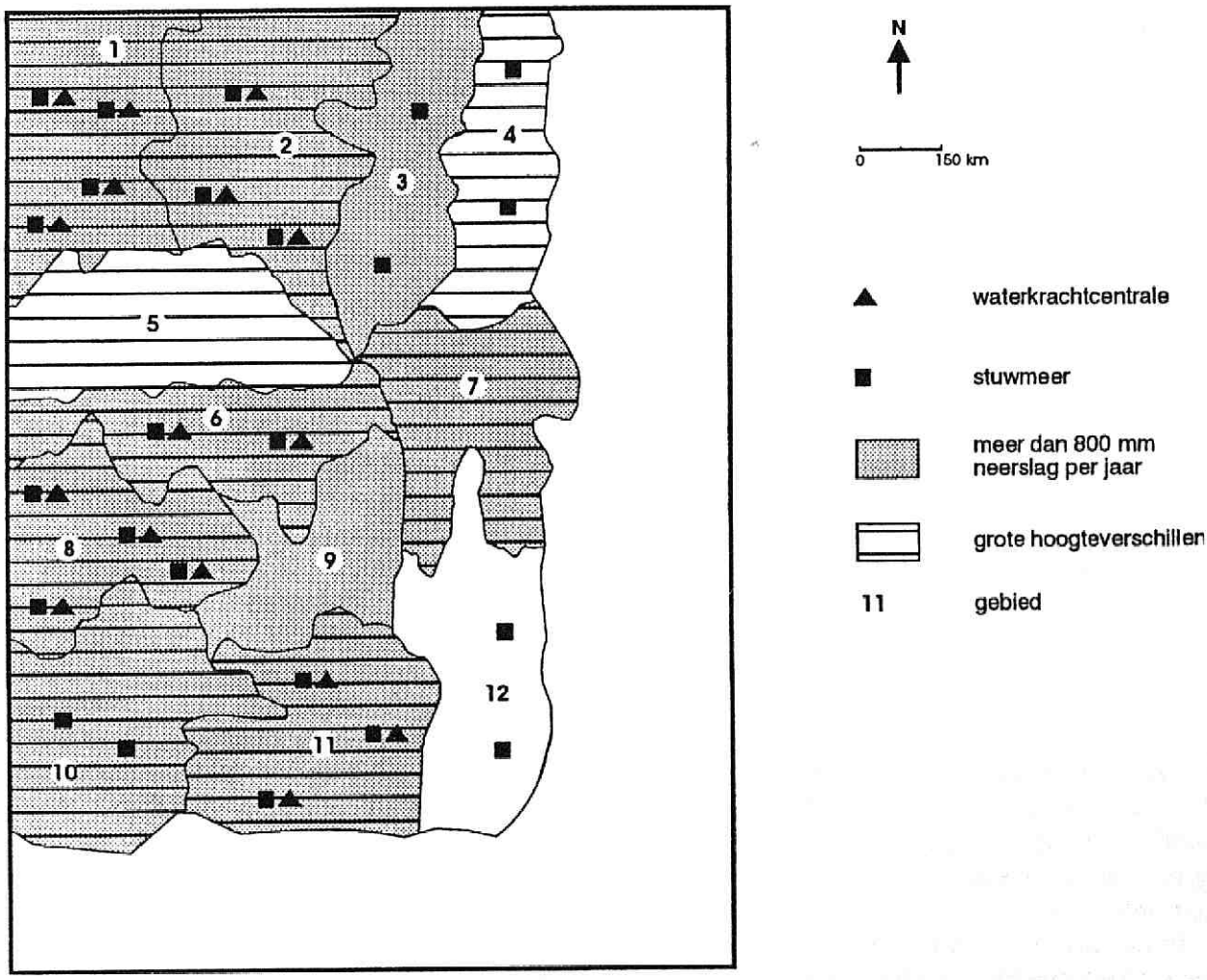
KAARTVAARDIGHEDEN	*DENKOPERATIES*
KAARTLEZEN	IDENTIFICEREN
KAARTANALYSE	CLASSIFICEREN & RELATEREN
KAARTINTERPRETATIE	INTERPRETEREN

Schema 1. Typen kaartvaardigheden.

geledingen en ruimtelijke samenhangen op een kaart". Kaartanalyse wordt onderverdeeld in twee denkoperaties: classificeren en relateren. Classificeren betreft het groeperen en/of vergelijken van (verzamelingen van) objecten. Relateren betreft het leggen van relaties tussen (verzamelingen van) objecten op een kaart.

Voorbeelden van identificeren, classificeren en relateren zijn bij kaart 1 weergegeven door resp. vraag 1, vraag 2 en vraag 3. Kaartinterpretatie is "het geven van verklaringen of voorspellingen aan de hand van de op de kaart ontdekte ruimtelijke samenhangen waarbij ook andere informatiebronnen dan de kaart worden geraadpleegd" (Van der Schee, 1992). In het hier beschreven experiment wordt alleen uitgegaan van informatie die reeds door de kaart zelf wordt gegeven om zodoende het aantal variabelen in het experiment controleerbaar te houden. Kaartinterpretatie is in het experiment derhalve buiten beschouwing gelaten.

Vraag 1: Noem de kenmerken van gebied 10.
Vraag 2: Noem alles wat de gebieden 6, 8 en 10 gemeenschappelijk hebben.
Vraag 3: Je krijgt opdracht om in één gebied nieuwe waterkrachtcentrales te bouwen.
Kies uit de gebieden waar nog geen waterkrachtcentrale is.
Ik ga bouwen in gebied nummer: ....
Omdat:
1 .....
2 .....
3 .....



Kaart 1. De lokatie van waterkrachtcentrales.

Uit eerder onderzoek (Van Dijk et al., 1994; Van der Schee et al., 1992,1994) is gebleken dat brugklasleerlingen vrij goed in staat zijn tot het identificeren van objecten. Meer moeite hebben ze met het classificeren van objecten of verzamelingen van objecten. Het leggen van relaties op een kaart geeft de meeste problemen.

Daarnaast is uit het genoemde onderzoek gebleken dat het soort kaarten waarmee gewerkt wordt invloed heeft op de complexiteit van kaartvraagstukken. Kaarten met puntsymbolen, vlaksymbolen en lijnen die beweging aangeven worden resp. spreidingskaarten, geledingskaarten en interaktiekaarten genoemd. In spreidings- en geledingskaarten wordt een (verticaal) verband tussen verschijnselen in een bepaald gebied getoond (zie kaart 1). Interaktiekaarten laten een (horizontaal) verband zien waarbij afstandsverval een rol speelt. Een voorbeeld hiervan is de afname van het aantal busritten tussen steden van gelijke grootte naarmate de steden verder uit elkaar liggen. Op interaktiekaarten wordt door leerlingen uit het voortgezet onderwijs in het algemeen slechter gescoord dan op spreidings- en geledingskaarten (Van der Schee, 1987, p.231; Van der Schee et al., 1992, 1994; Van Dijk et al., 1994).

Tenslotte is gebleken dat naarmate het aantal spreidingen en/of geledingen op een kaart toeneemt, de kaartvaardigheidsscores van leerlingen op zowel relateren, classificeren als identificeren afnemen. Als bijvoorbeeld één van de geledingen op kaart 1 zou worden weggelaten, dan kunnen meer leerlingen een samenhang op de kaart ontdekken.

Hierboven zijn drie aspecten genoemd die de complexiteit van een kaartvraagstuk in het huidige experiment beïnvloeden. Dit betreft ten eerste de denkoperatie die wordt gevraagd (identificeren, classificeren of relateren), ten tweede het soort kaarten waarmee wordt gewerkt (spreidings- en geledingskaarten of interaktiekaarten) en ten derde het aantal spreidingen en/of geledingen dat op een kaart staat weergegeven.

De vraag is nu of metacognitieve handelingen met behulp van computerlesmateriaal het leren van kaartvaardigheden, en met name het relateren van in kaart gebrachte informatie, kan vergemakkelijken. Anders gezegd: is computer-lesmateriaal met metacognitieve hulp effectiever (in de zin van meer leereffect op het gebied van kaartvaardigheid) dan computer- lesmateriaal zonder metacognitieve hulp? Daartoe moeten metacognitieve handelingen worden 'vertaald' naar de vakspecifieke situatie.

Collins et al. (1989) pleiten voor een systeem dat "cognitive apprenticeship" wordt genoemd. In dit systeem worden leerlingen aan de hand van verschillende methodes ("modeling", "coaching", "scaffolding", "fading", "articulation", "reflection" en "exploration") geholpen om cognitieve en metacognitieve vaardigheden te gebruiken om kennis te verwerven. De functie van de eerste drie methoden "modeling", "coaching" en "scaffolding" beschrijven Collins et al. als: "to help students acquire an integrated set of cognitive and metacognitive skills through processes of observation and of guided and supported practice" (p.481). Door middel van "modeling" wordt eerst aan de leerlingen getoond hoe een expert een bepaald probleem oplost. Vervolgens mag de leerling zelf soortgelijke problemen oplossen voorzien van feedback ("coaching") waarbij de (te) moeilijke onderdelen van een vraagstuk door de docent worden verricht ("scaffolding"). Naarmate de leerling meer taken zelfstandig kan oplossen neemt de sturende en helpende rol van de docent af ("fading"). De functies van "articulation" en "reflection" formuleren Collins et al. als volgt: "to help students both focus their observations of expert problem solving and gain conscious access to (and control of) their own problem-solving strategies" (p.481). De leerling wordt regelmatig gevraagd om over zijn of haar oplossingsmethode na te denken en deze toe te lichten ("articulation"). Met "reflection" wordt bedoeld dat de leerling zijn of haar eigen probleem-oplos methode vergelijkt met die van andere studenten, een expert en/of een intern cognitief model van expertise. Tenslotte raden Collins et al. aan om via "exploration" de autonomie van de lerende te vergroten, dit impliceert dat de leerling op nieuwe terreinen problemen gaat oplossen.

In het hier beschreven experiment zijn twee computer-lesmodulen (één met en één zonder metacognitieve hulp) ontworpen en getest. De module met metacognitieve hulp is geïnspireerd op het systeem dat Collins et al. (1989) aanraden. In deze methode zijn alle fasen van het model

"cognitive apprenticeship" verwerkt met uitzondering van "exploration" omdat deze moeilijk te realiseren valt zonder de beide modulen al te veel in tijdsbesteding uiteen te laten lopen. Verondersteld wordt dat de kaartvaardigheid van leerlingen positief wordt beïnvloed als leerlingen met behulp van de door Collins et al. aangeraden methoden bewust worden gemaakt van de stappen die ze zelf nemen en de stappen die ze zouden moeten nemen om een kaartprobleem op te lossen. Omdat deze metacognitieve activiteiten een vakspecifieke invulling kregen is de naam strategische hulp gebruikt.

De module zonder strategische hulp laat de leerlingen kaartvaardigheden uitvoeren zonder over een bepaalde 'aanpak' te spreken. In deze module krijgen de lerenden vragen voorgelegd zonder dat hen duidelijk wordt gemaakt welk verband er tussen de vragen bestaat. Evenmin wordt expliciet aangegeven welke activiteiten de lerende moet ondernemen als een probleem niet opgelost kan worden. De leerling wordt in deze module dus geacht zelf te bedenken welke cognitieve vaardigheden nodig zijn om een vraagstuk op te lossen.

Hypothese en vraagstellingen

In het huidige experiment is de volgende hoofdhypothese is geformuleerd: computer- lesmateriaal voorzien van strategische hulp levert meer leereffecten op dan hetzelfde lesmateriaal waar deze hulp niet aan is toegevoegd.

Een interessante vraag hierbij is in hoeverre relatief bekwame leerlingen in kaartvaardigheden meer of minder baat hebben bij strategische hulp dan minder bekwame leerlingen. Temeer daar is gebleken dat succesvolle leerlingen vaker regulatie-activiteiten hanteren dan minder succesvolle leerlingen (De Jong, 1992, p.180). Volgens Van Hout-Wolters (1992, p.17) zijn zelfregulatie-strategieën zelfs het allerbelangrijkste voor het behalen van goede leerresultaten.

Tenslotte zal in het onderzoek worden gekeken in hoeverre respectievelijk de factor geslacht en tijd, invloed hebben op de natoetscijfers van de leerlingen. Uit verschillend onderzoek is gebleken dat mannelijke adolescenten en volwassenen over het algemeen beter presteren op taken die visueel ruimtelijk inzicht vereisen dan vrouwelijke adolescenten en volwassenen. "The male advantage on spatial tests increases through the high school years up to a level of about .40 of a standard deviation" (Maccoby & Jacklin 1974, p.351). In een onderzoek van Van Dijk en Riezebos (1991) onder 600 brugklas en derde klas leerlingen werd een verband gevonden tussen sexe en aardrijkskunde-rapportcijfers. Jongens scoorden significant hoger. Op het gebied van kaartlezen (van topografische kaarten) hebben Gilmartin (1986) en Chang & Antes (1987) een onderzoek onder resp. eerstejaars geografie- en psychologiestudenten uitgevoerd waaruit bleek dat de mannelijke studenten significant beter presteerden op de vaardigheid kaartlezen dan de vrouwelijke studenten. Ook in een vooronderzoek van het hier beschreven experiment werden sexegebonden verschillen gevonden. Op de vragen waarbij twee of meer verschijnselen aan elkaar gerelateerd moesten worden, werd door jongens significant hoger gescoord dan door meisjes (Van Dijk et. al., 1993). In dit experiment zal worden onderzocht of deze verschillen zich voordoen in zowel de voortoets als de natoets.

METHODE

Proefpersonen

Het onderzoek is uitgevoerd op een brede scholengemeenschap in Uden met brugklas- leerlingen van verschillende niveau's. Van deze school zijn twee willekeurige mavo/havo brugklassen en twee willekeurige havo/vwo brugklassen geselecteerd (in totaal 113 leerlingen).

Als gevolg van absentie van drie leerlingen bij de natoets zijn in totaal de gegevens van 58 mavo/havo leerlingen en 52 havo/vwo leerlingen verwerkt. Het betrof 53 meisjes en 57 jongens.

Onderzoeksopzet en procedure

Alle proefpersonen uit de mavo/havo en havo/vwo klassen die bij het experiment zijn betrokken zijn op basis van alfabetische volgorde in twee groepen verdeeld (zie schema 2) zodat twee

	VT	module 1 zonder str.hulp	module 2 met str.hulp	NT
groep 1	+	+		+
groep 2	+		+	+

Schema 2. Opzet van het experiment.

theoretisch gelijke groepen van ieder 55 leerlingen ontstonden. Groep 1 heeft deelgenomen aan een lesmodule zonder strategische hulp. Groep 2 kreeg wel strategische hulp voor en tijdens de lesmodule. De toewijzing van de twee condities aan groep 1 en groep 2 was willekeurig. In totaal hebben zowel groep 1 als groep 2 twee aparte lesuren aan het lesmateriaal besteed.

Er is bewust voor gekozen om mavo/havo groepen niet bij voorbaat als zwakke groep te betitelen en havo/vwo groepen als sterke groep. Aangezien het brugklassen betreft die zijn samengesteld op basis van adviezen van verschillende basisscholen, wordt verondersteld dat het prestatieniveau binnen zowel mavo/havo klassen als havo/vwo klassen vrij heterogeen zal zijn.

Alle leerlingen kregen een voortoets (VT), een computermodule van 2 lesuren en een natoets (NT) (zie schema 2). De natoets vond voor alle leerlingen 6 of 7 dagen na de laatste les plaats. Het aantal dagen tussen voortoets en natoets varieerde van minimaal 14 tot maximaal 18 dagen.

Lesmateriaal

Zoals vermeld zijn met het oog op de hoofdhypothese (computer-lesmateriaal voorzien van strategische hulp levert meer leereffecten op dan hetzelfde lesmateriaal waar deze hulp niet aan is toegevoegd) twee lesmodulen ontwikkeld. Beide modulen zijn vooral bedoeld om de vaardigheid in het relateren te verbeteren aangezien deze vaardigheid het minst ontwikkeld is bij brugklasleerlingen (Van der Schee et al.; 1992, 1994; Van Dijk et al., 1994).

Zowel module 1 als module 2 bestaat uit lesmateriaal waarin de stof, opgesplitst in (eenvoudiger) deelvragen, aan leerlingen voorgelegd wordt. Verondersteld wordt dat leerlingen die geen verbanden kunnen herkennen op een kaart, geholpen kunnen worden door hen eerst te laten classificeren en/of te laten identificeren. Aangezien in beide modulen, afhankelijk van de leerlingantwoorden, verschillende leerwegen mogelijk moeten zijn, leent het lesmateriaal zich er uitstekend voor om per computer te worden aangeboden in de vorm van geprogrammeerde instructie. Hieronder wordt eerst module 1 (zonder strategische hulp) beschreven. In de daarop volgende alinea komt module 2 (met strategische hulp) aan de orde.

Figuur 1 laat zien dat iedere leerling met dezelfde vraag start. Op basis van de samenhang die op de kaart getoond wordt moet de leerling een gebied kiezen waar een bepaald verschijnsel zich voor zou kunnen doen. Als deze vraag juist is beantwoord moet de leerling zijn of haar antwoord beargumenteren. Aan de hand van kernwoorden en synoniemen controleert het programma of de argumentatie juist is. (Spelfouten zijn mogelijk mits er niet meer dan 1 spelfout per 3 letters wordt gemaakt). De feedback na het eerste antwoord op 'relateren argumenteren' is beperkt tot "Dit antwoord is juist", "Dit antwoord is niet juist" of "Dit antwoord is niet volledig", afhankelijk van het gegeven antwoord. Bij een fout antwoord op de vraag ten aanzien van de gebiedskeuze of de argumentatie zal de leerling in module 1 naar een "classificatie-vraag" worden gestuurd. Bij deze vraag moeten leerlingen de gemeenschappelijke eigenschappen van verschillende gebieden noemen. Als deze vraag goed wordt beantwoord, dan krijgt de leerling weer de vraag ten aanzien van de gebiedskeuze (als die nog niet goed was beantwoord) of anders de vraag ten aanzien van de argumentatie van de gebiedskeuze. Na een fout op classificeren krijgt de leerling een "identificatie-vraag" voorgelegd. Hier moet de leerling alle kenmerken van één gebied herkennen. Op iedere vraag kan maximaal twee keer geantwoord worden. Als

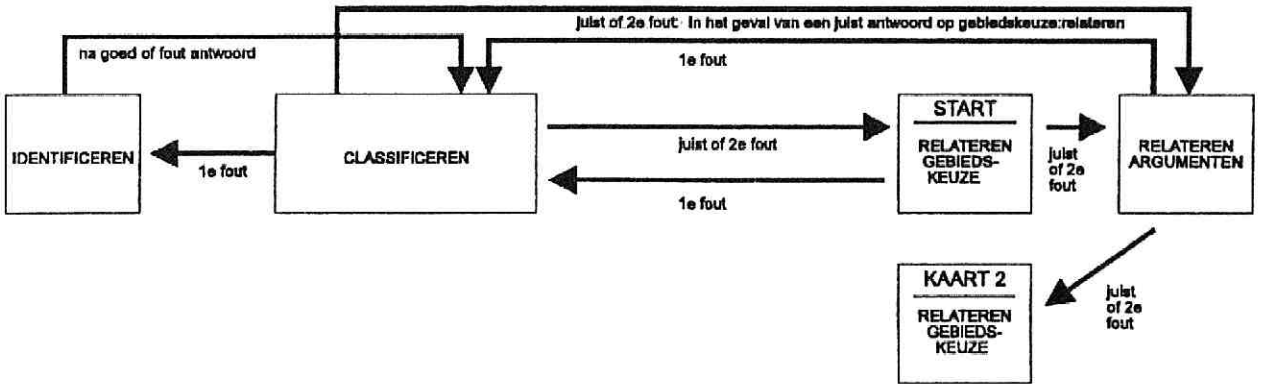


Fig. 1. Verwijzingsstructuur binnen module 1.

een vraag voor de tweede maal fout wordt beantwoord, wordt het juiste antwoord op het beeldscherm weergegeven. Na (maximaal 2 pogingen op) identificeren keert de leerling terug naar classificeren. Bij een juist antwoord of een tweede fout op classificeren wordt de leerling naar relateren (terug)gestuurd. Een leerling die een "relatie-vraag" goed beantwoordt of twee keer fout heeft beantwoord gaat door naar een volgende "relatie-vraag" bij een kaart waarop meer verschijnselen staan. Het lesmateriaal bestaat uit 12 kaarten. Alle antwoorden van leerlingen worden op schijf opgeslagen.

Een verkennend onderzoek in twee brugklassen met deze eerste module zonder strategische hulp toonde aan dat dit lesmateriaal met name op de denkoperatie relateren tot significant meer leerwinst leidde onder de experimentgroep dan onder de controlegroep die het lesmateriaal niet had ontvangen (Van Dijk et al, 1994). Desondanks bleef , gezien de succesvolle experimenten met lesmateriaal met regulerende activiteiten, het vermoeden bestaan dat een aantal leerlingen beter zouden kunnen.

In de tweede module die is voorzien van strategische hulp wordt eerst getoond hoe een expert aan de hand van vier stappen een kaartprobleem oplost ("modeling").
STAP 1) WAT staat er op de kaart?
STAP 2) WAT staat WAAR op de kaart?
STAP 3) Zie je een VERBAND op de kaart, zo ja WELK?
STAP 4) CONTROLEER of dit verband geldt voor ALLE GEBIEDEN in de kaart.
Hierna gaat de leerling met deze stappen oefenen aan de hand van een simpele kaart waarbij regelmatig feedback wordt gegeven ("coaching"). Vervolgens krijgt de leerling de mogelijkheid om zelfstandig een samenhang op een kaart te ontdekken. Als dit niet lukt worden de vier stappen en hun uitwerking op het scherm vertoond ("scaffolding"). Als de leerling wel zelfstandig de samenhang op een kaart kan ontdekken, worden de stappen en hun uitwerking niet gegeven ("fading").

Na deze kleine oefening begint de leerling aan het pakket kaartvaardigheden zoals beschreven bij figuur 1. Na een fout bij het relateren wordt de leerling strategische hulp geboden. De leerling wordt dan gevraagd na te denken over hoe hij of zij aan een antwoord is gekomen. Vervolgens volgt een scherm waarop de vier bovengenoemde essentiële stappen staan afgebeeld. Leerlingen moeten de antwoorden m.b.v. de vier beschreven stappen opschrijven ("articulation"). Op deze manier wordt de leerling aangezet om over de stappen na te denken. Het programma vraagt daarna of de leerling alle stappen heeft uitgevoerd ("reflection"). Na het beantwoorden van deze vraag krijgt de leerling een herkansing om het juiste gebied te kiezen of om een juiste argumentatie in te tikken. Daarna vervolgt dit pakket zijn weg zoals het eerder beschreven lesmateriaal (zie figuur 1).

Toetsmateriaal

Voor het meten van de leereffecten van de twee ontwikkelde lesmodulen is een toets ontwikkeld die als voortoets (VT) en natoets (NT) dienst doet. Met opzet is er voor gekozen om voor de voor- en natoets hetzelfde toetsmateriaal te gebruiken met het oog op een gelijke moeilijkheidsgraad van beide toetsen. De toets bestaat uit drie groepen vragen: relateren (A- vragen), classificeren (B-vragen) en identificeren (C-vragen). Bij de vragen zijn twee soorten kaarten opgenomen: spreidings- en geledingskaarten enerzijds en interaktiekaarten anderzijds. Op de kaarten staan meer of minder verschijnselen (spreidingen, geledingen en/of interakties) afgebeeld. De vragen 1, 2 en 3 bij kaart 1 zijn voorbeelden van resp. identificeren, classificeren en relateren uit de toets. De spreidings- en geledingskaarten hebben als thema de spreiding van waterkrachtcentrales. In de gemakkelijkste spreidings-geledingskaart uit de toets wordt alleen het verband tussen de aanwezigheid van neerslag en waterkrachtcentrales getoond. In de moeilijkste spreidings-geledingskaart (kaart 1) is het verband tussen waterkrachtcentrales, stuwmeren, neerslag en hoogteverschillen uitgebeeld. De interaktiekaarten hebben forensisme als thema. De meest complexe interaktiekaart laat het verband tussen intensiteit van forensenstromen naar een gemeente A of B, de nabijheid, het opleidingsniveau, de werkloosheid en het aantal inwoners zien. In de minst complexe interaktiekaart wordt alleen het verband tussen nabijheid en de interaktie (wel of geen forensenstroom) getoond. In deze minst complexe kaart is te zien dat forensen alleen uit relatief dichtbij gelegen buurgemeenten komen.

Om leereffecten te vermijden zijn de kaartvragen gerangschikt van complex naar minder complex. Dit betekent niet alleen dat A-vragen vooraf gaan aan B- en C-vragen maar ook dat binnen deze drie categorieën een ordening is aangebracht op basis van het aantal spreidingen/geledingen en/of interakties op de kaarten. Kaarten met bijvoorbeeld zowel een interaktie, spreiding als een geleding gaan vooraf aan kaarten met alleen een interaktie en een spreiding.

Het aantal vragen per onderdeel, gelijk aan de maximaal haalbare score (max) is in tabel 1 weergegeven. Voor het berekenen van scores zijn de leerlingantwoorden omgezet in goed/niet goed antwoorden.

RESULTATEN

Voortoets- en natoetsscores van groep 1 en groep 2

Voordat de natoetsscores (NT) zijn berekend zijn eerst de voortoetsscores van groep 1 (zonder strategische hulp) en groep 2 (met strategische hulp) vergeleken. Groep 2 scoorde op bijna alle onderdelen van de voortoets hoger dan groep 1. De verschillen waren echter niet significant ($P > .100$). Desondanks wordt hier kort ingegaan op mogelijke bronnen van verschil tussen groep 1 en groep 2.

Theoretisch zijn groep 1 en groep 2 even capabel want de leerlingen zijn immers alfabetisch gerangschikt. Aspecten die niet constant zijn geweest betreffen het tijdstip waarop de voortoets is afgenomen (groep 2 heeft de voortoets een week later gemaakt dan groep 1) en de docenten die de toets hebben afgenomen. De onderzoeker heeft in groep 1 alle voortoetsen afgenomen terwijl de voortoets van de leerlingen uit groep 2 door docenten van de school is afgenomen. Gezien de aard van de toetsvragen is het niet waarschijnlijk dat informatieuitwisseling over de voortoets de scores heeft beïnvloed. Het effect van de docent die de toets heeft afgenomen op de motivatie en het zelfvertrouwen van leerlingen heeft mogelijk wel enige invloed gehad op de voortoetsscores.

In tabel 1 zijn de gemiddelde scores van groep 1 (zonder strategische hulp) en groep 2 (met strategische hulp) op de natoets weergegeven. Tussen haakjes staan de standaarddeviaties (s.d.). Uit tabel 1 valt af te lezen dat groep 2 op de A-vragen (relateren) significant hoger scoorde dan groep 1 bij een betrouwbaarheidspercentage van 95%. Op het totaal aan A-vragen scoorde groep 2 gemiddeld 6.5 en groep 1 gemiddeld 5.3. Als wordt gecorrigeerd voor het bestaande verschil bij de voortoets tussen groep 1 en groep 2, blijft het verschil op de A-vragen significant

Tabel 1.Gemiddelde natoets scores van groep 1 en 2 (N=110).

	max.	groep 1	(s.d)	groep 2	(s.d.)	t–waarde
A: Relateren	12	5.3	(2.4)	6.5	(2.2)	–2.64*
waarvan:						
INTeraktie–vragen	6	1.1	(1.1)	1.8	(1.4)	–2.28*
SPreid/Gel.vragen	6	4.2	(1.8)	4.8	(1.3)	–2.02*
B: Classificeren	10	6.8	(1.8)	6.7	(1.9)	0.16
waarvan:						
INTeraktie–vragen	5	1.9	(1.7)	1.9	(1.7)	0.06
SPreid/Gel.vragen	5	4.8	(0.4)	4.8	(0.6)	0.36
C: Identificeren	10	6.9	(2.3)	7.5	(2.3)	–1.23
waarvan:						
INTeraktie–vragen	5	2.0	(2.3)	2.5	(2.3)	–1.27
SPreid/Gel.vragen	5	4.9	(0.3)	4.9	(0.3)	0.00

Noot: * $p< 0.05$

($F(1,107)=4.504;p<.05$). Op de afzonderlijke onderdelen van de A-vragen: interactie-vragen en spreidings/geledings-vragen verdwijnt de significantie echter: resp. ($F(1,107)=3.737;p=.056$) en ($F(1,107)=2.016;p=.159$).

De gevonden verschillen op de A-vragen vallen makkelijk te verklaren. Aangezien met name relateren een probleem vormt voor leerlingen is het niet vreemd dat daar de meeste leerwinst is opgetreden. Het verschil in effect tussen de modulen met en zonder strategische hulp is gevonden op de relatie-vragen waar de meeste leerwinst te maken viel. Opgemerkt moet echter worden dat de verschillen tussen groep 1 en groep 2 in natoetsscores niet bijzonder groot zijn.

Op basis van de gemiddelde scores op A-vragen uit de voortoets zijn alle leerlingen uit de onderzoekspopulatie ingedeeld in 'zwakke' leerlingen met een gemiddelde < 3.19 en 'sterke' leerlingen met een gemiddelde > 3.19.

Omdat alle leerlingen op basis van alfabetische volgorde in groep 1 of groep 2 zijn geplaatst, en niet op basis van matching van 'zwakke' en 'sterke' leerlingen, zijn deze leerlingen niet gelijk verdeeld over groep 1 en groep 2. Van de 'zwakke' leerlingen zaten er 32 in groep 1 en 26 in groep 2. Bij de 'sterke' leerlingen zaten de meeste leerlingen in groep 2 (29 versus 23).

Tabel 2 laat zien dat van de 'zwakke' leerlingen uit groep 2, de natoetsscore op relatie-vragen significant hoger was dan van de 'zwakke' leerlingen van groep 1. Voor 'sterke' leerlingen op de natoets lijkt de strategische hulp weinig uit te maken (zie tabel 2). Een vergelijking van de gemiddelde natoetsscores van sterke leerlingen uit groep 1 en groep 2 leverde geen significante verschillen op.

Uit variantie analyse blijkt echter geen significante interactie tussen de variabele groep (met en zonder strategische hulp) en bekwaamheid in kaartvaardigheid (zwak en sterk) ($F(1,106)=1.370$; $p>0.10$). De veronderstelling dat met name 'zwakke' leerlingen baat hebben bij strategische hulp is in dit experiment dan ook niet aangetoond.

Andere leerlingvariabelen

Tenslotte blijft de vraag over of de leerlingscores op de voor- en natoets beïnvloed zijn door andere leerlingkenmerken.

Tabel 2. Gemiddelde natoets scores van 'zwakke' en 'sterke' leerlingen van groep 1 en 2 op A-vragen (N=110).

	groep 1	(s.d.)	groep 2	(s.d.)	t-waarde
'Zwakke ll.'					
A: Relateren	4.4	(2.6)	5.9	(2.4)	-2.19*
waarvan:					
INTeraktie-vragen	0.9	(1.1)	1.6	(1.5)	-1.97
SPreid/Gel.vragen	3.5	(2.0)	4.3	(1.4)	-1.71
'Sterke ll.'					
A: Relateren	6.6	(1.6)	7.0	(1.8)	- 0.98
waarvan:					
INTeraktie-vragen	1.4	(1.1)	1.8	(1.3)	- 0.96
SPreid/Gel.vragen	5.1	(1.2)	5.3	(0.9)	- 0.49

Noot: * p< 0.05

Uit tabel 3 blijkt dat op de spreidings- en geledingskaarten van de A-vragen, jongens bij de voortoets significant beter scoren.

Uit de natoetsscores bleken geen significante verschillen tussen jongens en meisjes (noch voor de groep als geheel noch voor groep 1 en groep 2 afzonderlijk). Tussen de variabelen geslacht en bekwaamheid in kaartvaardigheid werden geen interacties gevonden ($F(1,102) = .139$; $p > 0.10$). Evenmin tussen geslacht en groep ($F(1,102) = .187$; $p > 0.10$).

Blijkbaar zijn de verschillen tussen jongens en meisjes niet van dien aard dat ze onoverbrugbaar zijn. Overeenkomstige conclusies werden getrokken door Van der Schee (1987, p.174) op basis van een onderzoek naar het effect van lesmateriaal op de kaartvaardigheid onder 4-vwo-klassen.

Tabel 3. Gemiddelde scores van meisjes en jongens op A-vragen van voor- en natoets (N=110).

	meisjes	(s.d.)	jongens	(s.d.)	t-waarde
Voortoets:					
A: Relateren	2.9	(1.9)	3.5	(1.8)	-1.77
waarvan:					
INTeraktie-vragen	0.5	(0.5)	0.4	(0.4)	0.73
SPreid/Gel.vragen	2.4	(1.7)	3.1	(1.8)	-2.09*
Natoets:					
A: Relateren	5.9	(2.3)	5.9	(2.4)	0.06
waarvan:					
INTeraktie-vragen	1.4	(1.3)	1.4	(1.3)	0.27
SPreid/Gel.vragen	4.5	(1.6)	4.5	(1.7)	-0.12

Noot: * p< 0.05

Tabel 4. Correlaties tijdsbesteding aan computerlessen en scores op kaartvaardigheidsnatoets (N = 96).

Scores op:	Computer–tijd
A: Relateren	.0093
waarvan:	
INTeraktie–vragen	.0817
SPreid/Gel.vragen	–.0508
B: Classificeren	.0356
C: Identificeren	.0618

Noot: 2–tailed Signif: * –.01 ** –.001

Een andere variabele die de scores op de kaartvaardigheid zou kunnen beïnvloeden is de hoeveelheid tijd die is besteed aan de voortoets, natoets en/of het lesmateriaal.Tussen de hoeveelheid tijd die is besteed aan de voor- of natoets en de scores op de kaarttoets is geen verband gevonden. Belangrijker dan een eventuele correlatie tussen de hoeveelheid tijd die is besteed aan het maken van de voor- of natoets en de kaarttoets-score, is een eventueel verband tussen de tijdsbesteding aan het lessenpakket en de scores op de natoets. Als er immers een positief verband bestaat tussen de hoogte van de scores op de natoets en de hoeveelheid tijd die is besteed aan het leren van kaartvaardigheden, dan is tijd wellicht een verklarende factor voor de leerwinst van leerlingen. Daar de computerlessen met strategische hulp gemiddeld meer tijd in beslag nemen dan computerlessen zonder hulp, zou dit conclusies ten aanzien van het afzonderlijke effect van strategische hulp op de natoetsscores onmogelijk kunnen maken.

Uit tabel 4 valt echter af te leiden dat de hoeveelheid tijd die leerlingen achter de computer hebben "geleerd" niet of nauwelijks correleert met de scores op de natoets. De tijd achter de computer is berekend op basis van de twee computerlessen samen. In de praktijk bleken leerlingen uit groep 2, gemiddeld 3 minuten korter over de eerste computerles te doen en gemiddeld 7 minuten langer over de tweede computerles. Dat leerlingen uit groep 2 (met strategische hulp) de eerste computerles sneller maakten valt te verklaren uit de 'modeling' voorbeelden die zij voorafgaand aan de computerlessen hebben gezien. Zij begrepen zodoende eerder wat de bedoeling was dan de 1e groep die zich de kaartvaardigheden meer door "trial and error" eigen moest maken. Bij de tweede (moeilijker) computerles maakten ook leerlingen uit groep 2 veel fouten. Daar zij gevraagd waren om bij een fout antwoord expliciet op te schrijven hoe zij een vraagstuk hadden opgelost, deden deze leerlingen gemiddeld langer over deze tweede computerles dan groep 1 die niets hoefde op te schrijven.

Op basis van bovenstaande gegevens mag worden aangenomen dat de variabelen geslacht en tijd geen onafhankelijke invloed hebben gehad op de uiteindelijke natoetsscores.

DISCUSSIE

In deze bijdrage zijn de leereffecten van twee computermodulen op de kaartvaardigheid van brugklasleerlingen met elkaar vergeleken. Op basis van literatuur ten aanzien van de effecten van metacognitieve hulp in een domeinspecifieke context (Perkins & Salomon, 1989), 'de gecombineerde benadering' (WRR-rapport, 1992), is verondersteld dat aardrijkskundig lesmateriaal voorzien van strategische hulp meer leereffect oplevert dan hetzelfde lesmateriaal waar deze hulp niet aan toegevoegd is. De strategische hulp werd vorm gegeven naar voorbeeld van een systeem dat "cognitive apprenticeship" wordt genoemd (Collins, Brown & Newman,1989).

Daarnaast werd de vraag gesteld welk lesmateriaal voor 'zwakke' en 'sterke' leerlingen (d.w.z. leerlingen die minder of meer dan gemiddeld scoorden op de voortoets "relatie- vragen")

het meeste leereffect oplevert. Uit verschillende literatuur blijkt dat metacognitieve vaardigheden samenhangen met leerprestaties (Wang, Haertel & Walberg, 1990, p.36; De Jong, 1992, p.180) terwijl deze metacognitieve vaardigheden niet geheel door intelligentie worden verklaard (Veenman & Elshout, 1992, p.299).

Wat de hypothese van dit experiment betreft, lijkt het huidige materiaal bescheiden aanwijzingen te geven dat een computermodule met strategische hulp meer leereffect oplevert dan een computermodule zonder strategische hulp. De verschillen in leereffect tussen beide modulen waren echter niet groot. Deze bescheiden resultaten moeten worden gezien in combinatie met het beperkte aantal lessen dat is besteed aan de beide modulen. Van Hout- Wolters (1992) merkte op dat "het leren van cognitieve strategieën, met name het leren van zelfregulatiestrategieën, niet zo snel gaat als we zouden willen. Het heeft meer tijd nodig dan een korte training (Van Hout-Wolters, 1992, p.21)". Daarnaast moet rekening worden gehouden met het eventuele effect van bepaalde eigenschappen van het gebruikte instructie-medium. De Jong (1992, p.188) noemde het gebrek aan affectieve componenten van de computer ten opzichte van een leerkracht. In dit lesmateriaal viel deze eigenschap op in die zin dat sommige leerlingen een opdracht om hun gedachten op papier te zetten negeerden of slechts zeer marginaal uitvoerden. Een docent zou in zo'n situatie doorvragen. Op basis van de uitkomsten van dit experiment zal met 'de gecombineerde benadering' opnieuw worden geëxperimenteerd, echter door middel van een iets langere training en op grotere schaal.

Ten aanzien van de vraagstelling in hoeverre relatief bekwame leerlingen in kaartvaardigheid meer of minder baat hebben bij strategische hulp dan minder bekwame leerlingen, is vastgesteld dat in dit experiment geen interaktie heeft plaatsgevonden tussen bekwaamheid in kaartvaardigheid en groep (met of zonder strategische hulp). De kanttekening moet hierbij echter worden gemaakt dat 'zwakke' en 'sterke' leerlingen in het hier beschreven experiment zijn geoperationaliseerd als bekwaamheid op de voortoets kaartvaardigheden. In hoeverre deze leerlingen reeds over metacognitieve vaardigheden beschikten is derhalve niet bekend. Het verzamelen van meer gedetailleerde informatie omtrent de reacties van 'zwakke' en 'sterke' leerlingen op de strategische hulp in toekomstig onderzoek zou meer kwalitatieve informatie kunnen opleveren over de effecten van strategische hulp op het leergedrag van deze leerlingen.

Tenslotte is in het experiment gezocht naar samenhang tussen het leerlingkenmerk geslacht en de factor tijd enerzijds en de natoetsscores van groep 1 en groep 2 anderzijds. De resultaten gaven geen aanleiding te veronderstellen dat deze variabelen een eigen invloed hadden op de onderzoeksresultaten.

LITERATUUR

Anderson, J.R. (1990). *Cognitive psychology and its implications.* (3rd ed). New York: W.H. Freeman and Company.

Boekaerts, M. & Simons, P.R.J. (1993). *Leren en instructie: Psychologie van de leerling en het leerproces.* Assen: Dekker & van de Vegt.

Bruner, J.S., Goodnow, J.J. & Austin, A. (1988). Categorization. In S.K. Reed (ed.). *Cognition, theory and applications* (p.152-174). Pacific Grove, California: Brooks/Cole Publishing Company.

Chang, K. & Antes, J.R. (1987). Sex and cultural differences in map reading. *The American Cartographer, 14,* 29-42.

Chase, W.G. & Simon, H.A. (1973). Perception of chess. *Cognitive Psychology, 4* , 55-81.

Collins, A. & Brown, J.S. (1988). The computer as a tool for learning through reflection. In H. Mandl & A. Lesgold (eds.). *Learning issues for intelligent tutoring systems* (p.1-18). New York: Springer Verlag.

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* (p.453-494). Hillsdale, NJ.: Erlbaum.

Dijk, H. van & Riezebos, A. (1991). Meisjes, jongens en schoolaardrijkskunde: grote verschillen in prestaties, waardering en interesses. *Geografisch Tijdschrift 25,* 71-75.

Dijk, H. van, Trimp, H. & Zijpp, T. van der (1993). *Kaartvaardigheden en Aardrijkskunde*. Amsterdam: CEG/VU: interne publikatie.

Dijk, H. van, Schee, J. van der, Trimp, H. & Zijpp, T. van der (1994). Map skills and geographical knowledge. *International Research in Geographic and Environmental Education, 3*, 68-80.

Flavell, J.H. (1979). Metacognition and cognitive monitoring. A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906-911.

Gilmartin, P.P. (1986). Maps, mental imagery and gender in the recall of geographical information. *The American Geographer, 13*, 335-344.

Goldstein, I. & Papert, S. (1977). Artificial Intelligence, language, and the study of knowledge. *Cognitive Science, 1*, 84-124.

Handboek Basisvorming voor de praktijk van het voortgezet onderwijs. Houten: Bohn Stafleu Van Loghum.

Hoekveld-Meijer, G. & Schutte, G.J. (1988). *Aardrijkskunde gebiedenderwijs*. NGS 82. Amsterdam: KNAG/VU.

Hout-Wolters, B.H.A.M. van (1992). *Cognitieve strategieën als onderwijsdoel*. Groningen: Wolters-Noordhoff.

Jong, F.P.C.M. de (1992). *Zelfstandig leren: Regulatie van het leerproces en leren reguleren: een procesbenadering*. Proefschrift, Tilburg: Katholieke Universiteit Brabant.

Maccoby, E.E. & C.N. Jacklin (1974). *The psychology of sex differences*. Stanford: Stanford University Press.

Muehrcke, P. (1978). *Map use: reading, analysis and interpretation*. Madison: J.P. Publications.

Olson, J.M. (1979). Cognitive cartographic experimentation. *The Canadian Cartographer, 16*, 34-144.

Ormeling, F.J. (1990). *Kartografie, visualisatie van ruimtelijke gegevens*. Delft: Universitaire Pers.

Palincsar, A.S. & Brown, A.L. (1984). Reciprocal teaching of comprehension fostering and comprehension monitoring activities. *Cognition and Instruction, 1*, 117-175.

Perkins, D. & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18*, 16-25.

Resnick, L.B. (1987). *Education and learning to think*. Washington D.C: National Academy Press.

Scardamalia, M., Bereiter, C. & Steinbach, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science, 8*, 173-190.

Schee, J.A. van der (1987). *Kijk op kaarten*. NGS 36. Amsterdam: KNAG/VU.

Schee, J.A. van der, Dijk, H. van & Westrhenen, J. van (1992). Geographical procedural knowledge and map skills. In H. Schrettenbrunner & J. van Westrhenen (eds.). *Empirical research and geography teaching*. Amsterdam: KNAG/VU.

Schee, J.A. van der, Zijpp, T. van der, Hoekveld-Meijer, G. & Westrhenen, H. van (1994). Map skills and geography teaching. In F.G. Brinkman, J.A. van der Schee & M.C. Schouten (eds.). *Curriculum research: different disciplines and common goals*. Amsterdam: VU.

Schoenfeld, A.H. (1987). *What's all the fuss about metacognition?* In A.H. Schoenfeld (ed.). *Cognitive science and mathematics education*. London: Lawrence Erlbaum Associates.

Simons, P.R.J. & Verschaffel, L. (1992). Transfer: onderzoek en onderwijs. *Tijdschrift voor Onderwijsresearch, 17*, 3-16.

Stuurgroep Profiel Tweede Fase Voortgezet Onderwijs (1994). *De Tweede Fase vernieuwt. Scharnier tussen basisvorming en hoger onderwijs, deel 2*. Den Haag: Porsius Creatief BV.

Veenman, V.J. & Elshout, J.J. (1992). Intelligentie en metacognitieve vaardigheden. *Tijdschrift voor Onderwijsresearch, 17*, 290-302.

Vermunt, J.D.H.M. (1992). *Leerstijlen en sturen van leerprocessen in het hoger onderwijs naar procesgerichte instructie in zelfstandig denken*. Proefschrift, Amsterdam/ Lisse: Swets & Zeitlinger B.V.

Wang, M.C., Haertel, G.D., & Walberg, H.J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research, 84*, 30-43.

Westrhenen, J. van (1992). *Een nascholingsproject gericht op de vakoverstijgende karakteristieken van de basisvorming*. Amsterdam: IDO/VU.

W.R.R. (1992). *Jeugd in ontwikkeling*. Den Haag: SDU uitgeverij.

Manuscript ontvangen
Definitieve versie ontvangen

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 4, pp. 332-340

Leerstijlenonderzoek gerepliceerd; de samenhang tussen leerstijlen en intelligentie

V.V. Busato, F.J. Prins, C. Hamaker en K.H. Visser
Vakgroep Psychonomie, Universiteit van Amsterdam

ABSTRACT

This replication study shows considerable evidence for the four learning styles as proposed by Jan Vermunt (1992). In a sample of 300 students at the University of Amsterdam, faculty of Psychology, as well as in a sample of 287 students, the meaning directed, the reproduction directed, the application directed and the undirected style were found. Somewhat lower, but still comparable stability coefficients as Vermunt reported were also found. Also, the relation of intelligence on the learning style was investigated. On the meaning directed learning style a significant, positive influence was found. On the undirected learning style, however, a significant, negative influence was found. The undirected learning style appeared to be a negative predictor for study succes.

INLEIDING

In 1992 publiceerde Vermunt in zijn dissertatie de resultaten van zijn onderzoek naar de manieren waarop studenten leren in het wetenschappelijk onderwijs, en hoe daarmee rekening zou kunnen worden gehouden in dat onderwijs. De manier waarop een student leert, noemt Vermunt een *leerstijl*: een samenhangend geheel van leeractiviteiten die studenten gewoonlijk ontplooien. Leerstijlen verschillen van elkaar op een vijftal aspecten: de wijze waarop studenten de studiestof cognitief verwerken, de wijze waarop deze verwerking wordt gereguleerd, de affectieve processen die zich bij het studeren voordoen, de mentale leermodellen die eraan ten grondslag liggen en de leeroriëntaties van de studenten. Vermunt kwam tot vier, volgens hem wezenlijk verschillende leerstijlen. Ten eerste een *ongerichte* leerstijl: studenten met deze leerstijl zouden veel problemen hebben met het verwerken van de stof, geen grote hoeveelheden aankunnen en moeite hebben met het bepalen wat belangrijk is. Daarnaast zijn er studenten met een *reproduktiegerichte* leerstijl: deze studenten proberen de leerstof uit colleges en studieboeken zo goed als mogelijk te onthouden om die op tentamens te kunnen reproduceren. Voorts zijn er studenten wier studiegedrag zich kenmerkt door een *betekenisgerichte* leerstijl: zij pogen bedoelingen van auteurs te achterhalen, bredere verbanden te leggen, een eigen visie te ontwikkelen. Kortom, zij vertonen het meest een academische werkhouding. Ten vierde kunnen studenten onderscheiden worden die zich bedienen van een *toepassingsgerichte* leerstijl: deze studenten betrekken de studiestof op actuele verschijnselen, zijn zich er zeer bewust van wat ze ermee kunnen. Om deze leerstijlen vast te kunnen stellen, ontwikkelde Vermunt een leerstijldiagnose-instrument, de Inventaris Leerstijlen (ILS). Dit instrument is een vragenlijst over studie-aanpak, -motieven en -opvattingen. Met deze vragenlijst kunnen de vier leerstijlen in een puntenscore worden uitgedrukt. Hoewel een student derhalve van elke leerstijl kenmerken kan vertonen, veronderstelt Vermunt dat één leerstijl overheerst; dat is de leerstijl waarop de hoogste puntenscore wordt behaald. Voor een meer inhoudelijke bespreking van deze theorie en vragenlijst wordt verwezen naar Vermunt (1992).

Binnen een onderzoek naar de effectiviteit van de studievaardigheidscursus 'Begeleid Studeren '93' (Busato & Prins, 1993) voor eerstejaars studenten psychologie aan de Universiteit van

Adres: Universiteit van Amsterdam, Faculteit Psychologie, Vakgroep Psychonomie, Roetersstraat 15, 1018 WB Amsterdam.

Amsterdam, is een replicatie uitgevoerd van het leerstijlen-onderzoek van Vermunt (1992). Vraagstelling voor dit onderzoek is: worden dezelfde vier leerstijlen in Amsterdam gevonden? De ILS wordt daartoe bij eerstejaars psychologiestudenten afgenomen, zowel aan het begin van het studiejaar als halverwege. Voor het replicatie-onderzoek zijn de factor-analyse-procedures van Vermunt exact nagedaan met behulp van SPSS. Tevens wordt nagegaan hoe stabiel de leerstijlen blijven in vergelijking met wat Vermunt rapporteert. Daartoe wordt dezelfde correlatie-procedure van Vermunt toegepast. Verwacht wordt dat in Amsterdam, aan de faculteit der psychologie, dezelfde vier leerstijlen worden aangetroffen. Ook worden soortgelijke stabiliteitscoëfficiënten verwacht.

Een onzes inziens belangrijke variabele die Vermunt niet in zijn onderzoek heeft meegenomen, is intelligentie. Bekend is dat intelligentie een belangrijke voorspeller is voor leerprestaties (zie bijvoorbeeld Elshout, 1985; Elshout-Mohr, 1992; Elshout & Veenman, 1992; Veenman, Elshout & Busato, 1994). Riemersma (1993) noemt het in zijn boekbespreking van Vermunt welhaast noodzakelijk de variabele intelligentie te betrekken wanneer het om cognitieve verwerkingsactiviteiten gaat. Over de invloed die intelligentie heeft op de gehanteerde leerstijl in het wetenschappelijk onderwijs is echter niets bekend. Wij vermoeden dat deze variabele wel degelijk van invloed is op de gehanteerde leerstijl en dat intelligentie een belangrijk gedeelte van variantie daarin kan verklaren. Ook zal worden gekeken wat de specifieke bijdrage is van de leerstijl aan het studiesucces, gedefinieërd als het aantal behaalde studiepunten. Dit zal worden onderzocht op de manier waarop Veenman (1993) dat met de relatie tussen aanpak, intelligentie en leerprestaties deed. Het intelligentie-aspect voegen wij aan de replicatie toe.

METHODE

Proefpersonen

Proefpersonen waren eerstejaars psychologiestudenten van de Universiteit van Amsterdam. In totaal deden tijdens de eerste afname van de ILS 300 proefpersonen mee, de tweede afname werd uitgevoerd met 287 proefpersonen. Aan de studenten van de tweede steekproef werd toestemming gevraagd hun intelligentie-gegevens uit de testweek beschikbaar te stellen. De testweek is verplicht voor alle eerstejaars psychologie-studenten aan de UvA. In deze week wordt een grote verscheidenheid aan psychologische tests afgenomen. Van 247 proefpersonen zijn uiteindelijk de volledige gegevens beschikbaar.

Leerstijlen

De leerstijlen worden vastgesteld aan de hand van de ILS. Deze vragenlijst bestaat uit twintig subschalen die verschillende verwerkingstrategieën, regulatiestrategieën, mentale leermodellen en leeroriëntaties meten. De ILS telt 120 uitspraken. Van elke uitspraak dient een student op een vijfpuntsschaal aan te geven in hoeverre deze van toepassing is op het eigen studiegedrag. Afhankelijk van de formulering van de uitspraak variëren de antwoordmogelijkheden van 1 'ik doe dit zelden of nooit' tot en met 5 'ik doe dit vrijwel altijd', of van 1 'helemaal mee oneens' tot en met 5 'helemaal mee eens'. Een voorbeeld van een uitspraak, gericht op de verwerkingsstrategie 'relateren & structureren', is: *"Ik probeer studiestof die nieuw voor me is in verband te brengen met kennis die ik al heb over het behandelde onderwerp."* Een voorbeeld van een uitspraak, gericht op de regulatiestrategie 'stuurloos leergedrag', luidt: *"Ik realiseer me, dat het niet duidelijk is wat ik moet onthouden en wat niet."* Een voorbeeld van een uitspraak, gericht op het mentale leermodel 'stimulerend onderwijs', is: *"De docent moet mij stimuleren om zelf losse cursusonderdelen samen te voegen tot een geheel."* Een voorbeeld van een uitspraak, gericht op de leeroriëntatie 'beroepsgericht' luidt: *"Mijn belangrijkste doel bij deze studie is om mezelf voor te bereiden op de uitoefening van een beroep."* Vermunt (1992) rapporteert redelijke tot zeer goede betrouwbaarheden van de schalen van de ILS.

Tijdens verplichte werkgroepen, behorend bij het vak Inleiding in de Psychologie, werd de ILS voor de eerste keer afgenomen. Deze meting vond plaats in oktober, vlak na het eerste

tentamen. Om de stabiliteit van de leerstijlen te onderzoeken werd de ILS nogmaals afgenomen. Deze tweede afname vond plaats tijdens de laatste zitting in maart van dezelfde werkgroepen. Zodoende spreken we van een One-Group Pretest-Posttest Design (Judd, Smith en Kidder, 1991).

Intelligentie
Aan het begin van het studiejaar werd in het kader van de verplichte testweek de intelligentie vastgesteld aan de hand van zes tests: 'Conclusies III', 'Verbale Analogieën', 'Rekensnelheid', 'Verborgen Figuren', 'Reeksen' en 'Vocabulaire'. Alle tests vertegenwoordigen een component uit het 'Structure of Intellect Model' van Guilford (1967). 'Conclusies III' meet Cognitie van Semantische Systemen. 'Verbale Analogieën' meet Cognitie van Semantische Relaties. 'Rekensnelheid' meet Convergente productie van Symbolische Implicaties. 'Verborgen Figuren' meet Convergente productie van Figurale Transformaties. 'Reeksen' meet Cognitie van Symbolische Systemen. 'Vocabulaire' tenslotte meet de Cognitie van Semantische Eenheden. De gesommeerde scores op deze tests zouden een goede operationalisatie zijn van intelligentie (Guilford, 1980). Bovendien sluit een dergelijke operationalisatie aan bij de opvatting intelligentie als relatief te definiëren en niet als een elementaire intellectuele vaardigheid (bijvoorbeeld Elshout, 1976; Veenman, 1993).

Studiesucces
De propaedeuse psychologie aan de Universiteit van Amsterdam bestaat uit 42 studiepunten. Studiesucces werd geoperationaliseerd als het aantal behaalde punten in een heel studiejaar, dat liep van 1 september 1993 tot en met 31 augustus 1994.

RESULTATEN

In Tabel 1 en 2 worden de resultaten van de factor-analyses met betrekking tot de vier leerstijlen weergegeven. Deze corresponderen met de tabellen 5.7 en 5.8 op de bladzijden 116 en 117 van Vermunt (1992). Tabel 3 betreft de stabiliteit van de ILS-schalen en komt overeen met tabel 5.14 op bladzijde 126. In de tabellen 4 en 5 wordt de invloed van intelligentie op de ILS-schalen en de leerstijlen toegelicht.

Leerstijlen
In tabel 1 staan de factorladingen van de ILS-schalen in een 4-factoren Varimax-oplossing voor de gehele steekproef bij de eerste meting.

De factorladingen komen sterk overeen met de ladingen die Vermunt rapporteert, in een aantal gevallen tot op vrijwel de decimaal nauwkeurig. Op de eerste factor hebben de schalen relateren en structureren, kritisch verwerken, concrete verwerking, zelfsturing leerproces en -resultaten, zelfsturing leerinhoud en opbouw van kennis een hoge lading. Vermunt rapporteert ladingen van respectievelijk .71, .74, .59, .78, .69 en .73. Op de tweede factor laden de schalen memoriseren en herhalen, analyseren, externe sturing leerproces en externe sturing leerresultaten hoog. Vermunt rapporteert ladingen van respectievelijk .64, .69, .81 en .67. Op de derde factor laden de schalen stuurloos leergedrag, stimulerend onderwijs, samen studeren en ambivalent hoog. Vermunt rapporteert ladingen van respectievelijk .74, .60, .67 en .71. (Opgemerkt moet worden dat onze derde factor overeenkomt met de vierde factor in het onderzoek van Vermunt. Deze omdraaiing is te verklaren door kleine verschillen in eigenwaarden van de factoren, niet op grond van inhoudelijke verschillen.) Op de vierde factor tenslotte laden de schalen gebruik van kennis, certificaatgericht en beroepsgericht. Vermunt rapporteert ladingen van respectievelijk .68, .61 en .84. Bij ons laadt de schaal testgericht op deze factor (.53), dat is bij Vermunt niet het geval in de steekproef met Open Universiteit studenten maar wel in de steekproef met de Tilburgse studenten (.30). Factor 1 kan men interpreteren als de betekenisgerichte leerstijl, factor 2 als de reproductie-gerichte, factor 3 als de ongerichte en factor 4 als de

Tabel 1. Factorladingen van ILS-schalen in een 4-factoren Varimax-oplossing over onderzoek UVA-1 voltijdstudenten (hoofdcomponentenanalyse; N=300; ladingen >-.25 en <.25 weggelaten).

ILS-schalen	F1	F2	F3	F4
Verwerkingsstrategieën				
Diepte-verwerking				
Relateren & Structureren	.79			
Kritisch verwerken	.74			
Stapsgewijze verwerking				
Memoriseren & Herhalen		.69		
Analyseren	.45	.55		
Concrete verwerking	.58			
Regulatiestrategieën				
Zelfsturing				
Leerproces & -resultaten	.80			
Leerinhoud	.70			
Externe sturing				
Leerproces		.75		
Leerresultaten		.63		
Stuurloos leergedrag			.75	
Mentale Leermodellen				
Opbouw van kennis	.68			
Opname van kennis	-.25	.57	.33	.38
Gebruik van kennis				.63
Stimulerend onderwijs			.63	
Samen studeren			.61	
Leeroriëntaties				
Persoonlijk geïnteresseerd	.32			
Certificaatgericht	-.29	.40		.61
Testgericht			.31	.53
Beroepsgericht				.74
Ambivalent	-.31		.62	
Eigenwaarde	3.92	3.34	1.71	1.25
% verklaarde variantie	19.6	16.7	8.5	6.3
Cumulatief %	19.6	36.3	44.8	51.1

toepassingsgerichte leerstijl. De door de vier factoren verklaarde variantie ligt vrijwel even hoog als bij Vermunt (51.1 % in Amsterdam tegenover 54.6 % voor de OU en 52.5 % voor de KUB).

In tabel 2 staan de factorladingen van de ILS-schalen in een 4-factoren Varimax-oplossing voor de gehele steekproef bij de tweede meting.

Er wordt wederom een sterke overeenkomst gevonden met Vermunt, als ook met de factorladingen van onze eerste meting. De eerste factor laat zich weer interpreteren als de betekenisgerichte; de schalen relateren en structureren, kritisch verwerken, concrete verwerking, zelfsturing leerproces en -resultaten, zelfsturing leerinhoud en opbouw van kennis laden hoog op deze factor. In de tweede factor herkennen we de reproductie-gerichte leerstijl; de schalen memoriseren en herhalen, analyseren, externe sturing leerproces, externe sturing leerresultaten laden hoog. De derde factor laat zich wederom interpreteren als de ongerichte leerstijl; de kenmerkende schalen stuurloos leergedrag, stimulerend onderwijs, samen studeren en een ambivalente leeroriëntatie laden hoog op deze factor. De vierde factor kan weer worden geïnterpreteerd als

Tabel 2. (Tweede meting) Factorladingen van ILS-schalen in een 4-factoren Varimax-oplossing over onderzoek UVA-1 (hoofdcomponentenanalyse; N=287; ladingen >-.25 en <.25 weggelaten).

ILS-schalen	F1	F2	F3	F4
Verwerkingsstrategieën				
Diepte-verwerking				
Relateren & Structureren	.80			
Kritisch verwerken	.73			
Stapsgewijze verwerking				
Memoriseren & Herhalen		.66		.29
Analyseren	.39	.52		
Concrete verwerking	.66			
Regulatiestrategieën				
Zelfsturing				
Leerproces & -resultaten	.73			
Leerinhoud	.70			
Externe sturing				
Leerproces		.78		
Leerresultaten		.65		
Stuurloos leergedrag			.77	
Mentale Leermodellen				
Opbouw van kennis	.65			.37
Opname van kennis	-.35	.50	.35	.47
Gebruik van kennis	.27			.71
Stimulerend onderwijs			.48	.26
Samen studeren			.65	
Leeroriëntaties				
Persoonlijk geïnteresseerd	.48		-.27	
Certificaatgericht	-.39	.33		.53
Testgericht				.49
Beroepsgericht				.68
Ambivalent			.78	
Eigenwaarde	4.00	3.61	1.67	1.31
% verklaarde variantie	20.0	18.0	8.4	6.5
Cumulatief %	20.0	38.0	46.4	52.9

de toepassingsgerichte leerstijl; de schalen gebruik van kennis, certificaatgericht, testgericht en beroepsgericht laden hoog op deze factor. Het percentage door de vier factoren verklaarde variantie is eveneens vergelijkbaar: 52,9 %.

Stabiliteit

Net als Vermunt deed bij de steekproef van de OU-studenten (zie pagina 126 van zijn boek), hebben wij een stabiliteitsmeting verricht met een tijdsinterval van 20 weken (zie tabel 3). Dat is zesenhalve week langer dan het interval van Vermunt.

De correlaties van de verwerkingsstrategieën zijn van vergelijkbare grootte. Vermunt rapporteert respectievelijke (zie de volgorde in tabel 3) correlaties van .58, .55, .71, .67 en .58. De correlaties van de regulatie-strategieën, mentale leermodellen en leeroriëntaties liggen lager. Voor de regulatie-strategieën rapporteert Vermunt correlaties van .66, .69, .79, .61 en .68. Voor de mentale leermodellen rapporteert Vermunt correlaties van .72, .79, .70, .76 en .79. Voor de leeroriëntaties rapporteert Vermunt correlaties van .74, .72, .78, .80 en .74. Er is sprake van een redelijk grote stabiliteit in de scores op de ILS-schalen.

Tabel 3. Test-hertest correlaties (r) bij Uva-studenten van ILS-schalen met een interval van 20 weken (N=247).

ILS-schaal	r	ILS-schaal	r
Verwerkingsstrategieën		*Regulatiestrategieën*	
Diepte-verwerking		Zelfsturing	
Relateren & Structureren	.70**	Leerproces & -resultaten	.60**
Kritisch verwerken	.63**	Leerinhoud	.60**
Stapsgewijze verwerking		Externe sturing	
Memoriseren & Herhalen	.68**	Leerproces	.58**
Analyseren	.51**	Leerresultaten	.50**
Concrete verwerking	.61**	Stuurloos leergedrag	.57**
Mentale Leermodellen		*Leeroriëntaties*	
Opbouw van kennis	.54**	Persoonlijk geïnteresseerd	.54**
Opname van kennis	.73**	Certificaatgericht	.63**
Gebruik van kennis	.57**	Testgericht	.73**
Stimulerend onderwijs	.65**	Beroepsgericht	.68**
Samen studeren	.72**	Ambivalent	.61**

** : p<.01

Tabel 4. Correlaties (r) bij Uva-studenten van ILS-schalen met Intelligentie-scores (Intel.) (N=247).

ILS-schaal	Intel.	ILS-schaal	Intel.
Verwerkingsstrategieën		*Regulatiestrategieën*	
Diepte-verwerking		Zelfsturing	
Relateren & Structureren	.27**	Leerproces & -resultaten	.06
Kritisch verwerken	.26**	Leerinhoud	.14*
Stapsgewijze verwerking		Externe sturing	
Memoriseren & Herhalen	−.09	Leerproces	−.08
Analyseren	.04	Leerresultaten	.08
Concrete verwerking	.10	Stuurloos leergedrag	−.30**
Mentale Leermodellen		*Leeroriëntaties*	
Opbouw van kennis	.10	Persoonlijk geïnteresseerd	.11
Opname van kennis	−.24**	Certificaatgericht	−.14*
Gebruik van kennis	.06	Testgericht	−.16*
Stimulerend onderwijs	.01	Beroepsgericht	−.18**
Samen studeren	−.30**	Ambivalent	−.20**

* : p<.05; ** : p<.01

Intelligentie en leerstijlen

Aan de hand van de zes genoemde Guilford-tests die standaard worden afgenomen in de testweek, is een intelligentie-score bepaald. Voor de gehele steekproef gold: M=19.85, sd=4.7, N=392. In tabel 4 staan de correlaties met de afzonderlijke schalen van de ILS van de tweede meting.

Intelligentie verklaart tussen de acht en negen procent van de variantie in het gebruik van de strategieën relateren & structureren en kritisch verwerken. Van de regulatiestrategieën hangt zelfsturing leerinhoud positief samen met intelligentie. De regulatiestrategie stuurloos leergedrag hangt negatief samen met intelligentie. Het mentale leermodel opname van kennis hangt

Tabel 5. Correlaties (r) leerstijlen met intelligentie-score (Intel.) (N=247).

Leerstijl	Intel.
Betekenisgericht	.24**
Reproductiegericht	.05
Ongericht	-.27**
Toepassingsgericht	-.08

negatief samen met intelligentie. Alle leeroriëntaties behalve persoonlijke interesse hangen negatief samen met intelligentie.

In Tabel 5 staan de correlaties tussen de leerstijlen en intelligentie. De leerstijlen werden op basis van de resultaten van de factor-analyse als volgt geoperationaliseerd. Voor de betekenisgerichte leerstijl werden de scores op de strategieën relateren & structureren, kritisch verwerken, concrete verwerking, zelfsturing leerproces en leerresultaten, zelfsturing leerinhoud en het mentale leermodel opbouw van kennis bij elkaar opgeteld. De reproductiegerichte leerstijl werd berekend door de scores op de strategieën memoriseren, externe sturing leerproces en externe sturing leerresultaten op te tellen. De ongerichte leerstijl werd berekend door de scores op de strategieën stuurloos leergedrag, de mentale leermodellen stimulerend onderwijs en samen studeren en de leeroriëntatie ambivalent op te tellen. De toepassingsgerichte leerstijl werd berekend door de scores op het mentale leermodel gebruik van kennis en de leeroriëntatie beroepsgericht op te tellen. Opvallend is de significante en positieve correlatie tussen intelligentie en de betekenisgerichte leerstijl. Intelligentie draagt ruim vier procent bij aan de variantie. Intelligentie hangt daarentegen significant negatief samen met de ongerichte leerstijl. Ook dan verklaart intelligentie wederom ruim vier procent. Er wordt geen samenhang gevonden tussen de reproductiegerichte leerstijl en intelligentie. Datzelfde kan worden gezegd van het verband tussen de toepassingsgerichte leerstijl en intelligentie.

Intelligentie, leerstijl en studiesucces

Van 14 studenten was het aantal behaalde studiepunten niet bekend, deze correlationele analyses zijn derhalve uitgevoerd met 233 studenten. Het gemiddeld aantal behaalde studiepunten bedroeg 26.12 met een standaarddeviatie van 12.79. Er werd een significante, negatieve correlatie gevonden tussen de ongerichte leerstijl en het aantal behaalde studiepunten ($r=-.248$, $p<.01$). De partiële correlatie tussen de ongerichte leerstijl en het aantal behaalde studiepunten, met intelligentie uitgepartieërd, was eveneens significant ($r=-.211$, $p<.01$). De semipartiële correlatie tussen de ongerichte leerstijl en het aantal behaalde studiepunten, met intelligentie wél uitgepartieërd in de leerstijl maar níet uit het aantal behaalde studiepunten, was niet veel lager ($r=-.208$). Deze laatste correlatie is een iets zuiverdere schatting van de specifieke bijdrage die de gehanteerde leerstijl heeft op het aantal behaalde studiepunten. Men mag immers veronderstellen dat intelligentie ook een eigen inbreng heeft op het aantal behaalde studiepunten (zie voor een meer inhoudelijke bespreking van de semipartiële correlatie eventueel Veenman, 1993). Er werden geen opvallende ((semi-)partiële) correlaties tussen de andere leerstijlen en het aantal behaalde studiepunten gevonden. Opvallend was verder dat de betekenisgerichte leerstijl niet correleerde, ook niet negatief, met de ongerichte leerstijl, maar wel met de toepassingsgerichte ($r=.275$, $p<.01$) en de reproductiegerichte leerstijl ($r=.119$, $p<.05$). Dit kan mogelijk als illustratie worden gezien voor de overlap die er bestaat tussen de drie laatste leerstijlen en de relatieve autonome positie van de ongerichte leerstijl.

DISCUSSIE

Het is bekend dat weinig onderzoek binnen de psychologie gerepliceerd wordt. In dit onderzoek wordt echter aanzienlijke empirische steun gevonden voor de theorie van Vermunt. De gerepliceerde factor-analyses laten zien dat bij eerstejaars psychologie-studenten aan de Universiteit van Amsterdam dezelfde vier leerstijlen teruggevonden worden. Dit pleit voor het onderscheid in leerstijlen dat Vermunt maakt. Replicaties op grote schaal kunnen wellicht bijdragen aan een normering van de ILS, zodat de diagnostische waarde van het instrument kan worden vergroot.

Wat enigszins afwijkt in onze replicatie, is dat de correlaties tussen de eerste en tweede meting van de mentale leermodellen en leeroriëntaties lager uitvallen dan bij Vermunt. Dat zou kunnen komen doordat studenten in twintig weken een beter beeld van hun studie hebben gekregen dan in dertien en een halve week, het tijdsinterval dat Vermunt hanteerde. Omdat leermodellen en leeroriëntaties meer extern beïnvloed worden, lijkt deze verklaring plausibel. Maar deze zou bijvoorbeeld middels interviews nader kunnen worden onderzocht.

De variabele intelligentie blijkt een niet te verwaarlozen invloed te hebben in de variantie van de leerstijlen. Meest opvallend is de negatieve samenhang tussen intelligentie en de ongerichte leerstijl en de positieve samenhang tussen intelligentie en de betekenisgerichte leerstijl. Hoe minder intelligent een student, des te groter is de kans dat deze student een ongerichte leerstijl heeft. Hoe intelligenter een student, des te groter is de kans dat deze student een betekenisgerichte leerstijl heeft. Voor de toepassingsgerichte en reproductiegerichte leerstijl valt geen betekenisvolle samenhang te constateren met intelligentie. Het lijkt raadzaam om, wanneer de ILS als diagnostisch instrument vaker wordt toegepast, de factor intelligentie in ogenschouw te nemen.

Op grond van de correlationele analyses tussen intelligentie, leerstijl en studiesucces kan worden geconcludeerd dat de ongerichte leerstijl een negatieve voorspeller is van studiesucces, gedefinieërd als het aantal behaalde studiepunten in een jaar. Voor de drie andere leerstijlen kunnen wat dat betreft geen uitspraken worden gedaan. De significante correlaties tussen de betekenisgerichte, reproductiegerichte en de toepassingsgerichte leerstijl pleiten voor de notie van Vermunt (1992) dat iemand niet zozeer gekarakteriseerd kan worden door één leerstijl, maar door een combinatie.

Dit onderzoek kan worden gezien als een eerste aanzet tot onderzoek naar de relatie tussen intelligentie, leerstijl en studiesucces. Wellicht dat longitudinaal onderzoek meer gegevens op kan leveren over die relatie.

LITERATUUR

Busato, V.V. & Prins, F.J. (1993). *Begeleid Studeren '93.* Interne uitgave vakgroep psychonomie, Universiteit van Amsterdam.

Elshout, J.J. (1976). *Karakteristieke moeilijkheden in het denken.* Academisch proefschrift, Universiteit van Amsterdam.

Elshout, J.J. (1985). Een goed begin is het halve werk. Over determinanten van effectief beginnersgedrag. In F.J. Mönks & P. Span (eds.). *Hoogbegaafden in de samenleving* (p. 201-210). Nijmegen: Dekker & Van de Vegt.

Elshout-Mohr, M. (1992). Metacognitie van lerenden in onderwijsleerprocessen. *Tijdschrift voor Onderwijsresearch, 17,* 273-289.

Elshout, J.J. & Veenman, M.V.J. (1992). Relation between intellectual ability and working method. *Journal of Educational Research, 85,* 134-143.

Guilford, J.P. (1967). *The nature of human intelligence.* New-York: McGraw-Hill.

Guilford, J.P. (1980). Fluid and crystallized intelligences: two fanciful concepts. *Psychological Bulletin, 88(2),* 406-412.

Judd, C.M., Smith, E.R. & Kidder, L.H. (1991). *Research methods in social relations (sixth edition).* Fort Worth, TX: Holt, Rinehart and Winston.

Riemersma, F. (1993). Leerstijlen en procesgerichte instructie. *De Psycholoog, 9,* 367-368.

Veenman, M.V.J. (1993). *Intellectual ability and metacognitive skill. Determinants of discovery learning in computerized learning environments.* Academisch proefschrift, Universiteit van Amsterdam.

Veenman, M.V.J., Elshout, J.J., & Busato, V.V. (1994). Metacognitive mediation in learning with computer based simulations. *Computers in Human Behavior, 10(1)*, 93-106.

Vermunt, J.D.H.M. (1992). *Leerstijlen en sturen van leerprocessen in het hoger onderwijs.* Amsterdam/ Lisse: Swets & Zeitlinger.

Manuscript ontvangen 4-11-1994
Definitieve versie aanvaard 20-4-1995

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 4, pp. 341-362

Het relatieve belang van sociaal milieu en etnische herkomst voor de verklaring van onderwijsachterstanden*

Een overzicht van kwantitatief-empirische studies; bevindingen en problemen

G.W.J.M. Driessen
Instituut voor Toegepaste Sociale Wetenschappen (ITS), Universiteit Nijmegen

ABSTRACT

The educational position of many non-indigenous pupils in the Netherlands is very unfavourable. For some years now there has been an ongoing discussion about the question whether these disadvantages could have their origins in characteristics typical of migrants or whether they are in fact caused by the very same aspects that are also responsible for the low educational position of indigenous working-class children. Under discussion therefore is the relative importance of social background and ethnic origin. First, the results of some 30 studies in this field are described. The conclusion is that there are many similarities, but also major discrepancies in the findings. Next, therefore, possible explanations for these differences are looked into. Here, the emphasis is largely on factors relating to method. There appear to be various shortcomings and problems in this respect. In view of the fact, however, that these apply to virtually all of the studies, they can hardly help explain the differences found. It is recommended that the consequences of the various research approaches should be compared by means of simulation studies. Furthermore, this issue not only needs to be looked at methodologically, but should also be treated in greater depth conceptually.

1. INLEIDING

Dat allochtone kinderen het in het onderwijs (beduidend) slechter doen dan autochtone kinderen, daarover is men het in het algemeen wel eens (Driessen, 1990b). Aanzienlijk minder overeenstemming bestaat er echter over de achtergronden van dit verschil. In het bijzonder gaat de discussie die sinds een aantal jaren wordt gevoerd over het relatieve belang van twee gezinsstructurele kenmerken, te weten enerzijds het sociaal milieu waaruit de leerling afkomstig is en anderzijds de etnische herkomst van de leerling[1] (Driessen, 1990a; Van 't Hof & Dronkers, 1992).

Was de discussie lange tijd vooral een academische[2], inmiddels heeft zij – als gevolg van beleidskeuzes gebaseerd op een door Meijnen & Riemersma (1992) voor de Commissie Allochtone Leerlingen in het Onderwijs (1992) uitgevoerde literatuurstudie – ook praktische consequenties gekregen (MO&W, 1994). Tot voor kort was het namelijk zo dat de in het kader van het Onderwijsvoorrangsbeleid aan scholen toegekende extra faciliteiten voor een belangrijk deel werden bepaald door de etnische herkomst van de leerlingenpopulatie. Ofschoon de besluitvorming hierover nog niet is afgerond, ziet het er naar uit dat één van de milieu-indicatoren, het opleidingsniveau van de ouders, een zwaarwegender (of wellicht dominante) rol gaat spelen en de etnische herkomst wellicht helemaal verdwijnt als indicator.[3]

Hoewel sommigen misschien hoopten dat met de literatuurstudie van Meijnen & Riemersma een punt achter de milieu/herkomst-kwestie zou worden gezet, brandde in de daarna volgende

* Met dank aan J. Hutjes, G.W. Meijnen en twee anonieme beoordelaars voor hun commentaar op een eerdere versie van dit artikel.
Adres: Postbus 9048, 6500 KJ Nijmegen.

periode de discussie opnieuw los. Zo werd in 1994 tijdens de Zesde Sociaal-Wetenschappelijke Studiedagen een complete sessie aan dit onderwerp gewijd, waren er ook tijdens de Onderwijs Research Dagen 1994 papers die handelden over de verhouding milieu : herkomst, en was tijdens een in 1994 door NWO georganiseerd symposium een van de zwaartepunten 'De interrelatie van etnisch-culturele en sociaal-economische positie bij etnische minderheden'.

In de discussie kunnen grofweg twee kampen worden onderscheiden. Op de eerste plaats is er de aloude tegenstelling tussen aan de ene kant kwantitatief en aan de andere kant kwalitatief georiënteerde onderzoekers. De 'kwalitatieven' verwijten de 'kwantitatieven' dat zij met hun type onderzoek – juist vanwege het grootschalige en daaraan inherente oppervlakkige karakter – nooit tot de kern van de zaak kunnen doordringen. Met name zouden de processen, de mechanismen die ten grondslag liggen aan de statistische samenhangen niet blootgelegd kunnen worden. Vandaar dat zij een sterkere wisselwerking tussen kwantitatief en kwalitatief onderzoek bepleiten (vgl. Ledoux, 1994; Matthijssen & Teunissen, 1994; Pels & Veenman, 1994).

Echter, het kwalitatieven/kwantitatieven-debat vormt niet het enige thema; ook binnen de groep kwantitatieven bestaat onenigheid over de tot nu toe gepubliceerde onderzoeksbevindingen. Deze onenigheid vloeit niet alleen voort uit de soms afwijkende of zelfs tegenstrijdige bevindingen in de diverse onderzoeken, maar ook uit de interpretatie die daar vervolgens aan wordt gegeven. In dit artikel staat deze *kwantitatieve* benadering centraal.[4] Het doel van het artikel is een geactualiseerd beeld te schetsen van de onderhavige materie. Het eerste deel van het artikel bestaat uit een *inventarisatie en samenvatting* van onderzoeken waarin de verhouding milieu en herkomst aan de orde is gesteld. De invalshoek daarbij is in hoofdzaak beleidsgericht en wel in die zin, dat gezocht wordt naar een verantwoorde wijze waarop door de overheid middelen kunnen worden toegekend om een als ongewenst beschouwde situatie, i.c. de achterstandspositie van allochtonen, te verbeteren. Het gaat hier dus vooral om het zoeken naar indicatoren en criteria met behulp waarvan de overheid zodanig middelen kan verdelen, dat ze terecht komen bij die groepen die ze het hardst nodig hebben. Gelet op de aangetroffen diversiteit in onderzoeksbevindingen, wordt in het tweede deel van het artikel een aanzet gegeven voor de verklaring daarvan. In hoofdzaak gaat het daarbij om *methodologische aspecten*. Het artikel wordt ten slotte afgesloten met enkele discussiepunten.

2. ONDERZOEKSRESULTATEN

Tabel 1 geeft een overzicht van studies die zich min of meer expliciet hebben beziggehouden met het vraagstuk van de verhouding sociaal milieu en etnische herkomst in relatie tot onderwijsachterstanden. Bij het selecteren van de studies gold als eerste vereiste dat de uitgevoerde analyses kwantitatief van aard moesten zijn. Bovendien diende het te gaan om onderwijskundige en onderwijs-sociologische studies; dit betekent dat bijvoorbeeld specifiek taalkundig onderzoek niet in het overzicht is betrokken.

De beschrijving van de studies in de tabel vindt plaats aan de hand van een aantal categorieën. Deze worden hier slechts kort aangeduid; in een volgende paragraaf komen ze uitgebreider aan de orde. (1) *Bron* verwijst naar de publikatie waarin de studie staat beschreven. Soms komt het voor dat (nagenoeg) dezelfde analyses in meerdere publikaties worden beschreven; als dat zo is, dan wordt daar via '(vgl. ...)' naar verwezen. Behalve omwille van de volledigheid, zijn deze verwijzingen ook opgenomen vanwege accentverschillen in analyses en – in een enkel geval – conclusies. (2) *Data-verzameling* heeft betrekking op het jaar of de jaren (veelal: schooljaren) dat de dataverzameling heeft plaatsgevonden. (3) Onder *steekproef* worden drie aspecten vermeld: de steekproefgrootte, de onderwijsfase (niveau en leerjaar) waarop het onderzoek betrekking heeft, en het geografisch bereik van de steekproef (landelijk dan wel regionaal of lokaal). (4) De volgende rubriek betreft de wijze van *operationalisatie van milieu en herkomst*. (5) Dan volgen de *onderwijspositie-kenmerken* die zijn geanalyseerd. (6) De onder *analysetechniek/methode* vermelde varianten hebben alleen betrekking op de hier beschreven resultaten, dus niet op eventuele andere bevindingen in de betreffende publikaties. (7) De

belangrijkste *resultaten* worden vervolgens kort samengevat; zo mogelijk gebeurt dat in kwantitatieve termen. (8) De laatste kolom bevat de *conclusies* met het oog op de vraagstelling van dit artikel. Aangegeven wordt per analyse of milieu dan wel herkomst de meeste verklaringskracht bezit, of dat het relatieve gewicht gelijk is.[5] Tussen haakjes staat het aantal analyses, i.c. onderwijspositiekenmerken waarop deze conclusie betrekking heeft. Eventuele bijzonderheden ten slotte, worden in de vorm van noten onder de tabel vermeld.

In verband met de beknoptheid zijn in de tabel de volgende afkortingen gebruikt: N (aantal leerlingen in de steekproef); M (sociaal milieu), H (etnische herkomst); lo6 (lager onderwijs leerjaar 6), ba.o8 (basisonderwijs groep 8), vo1 (voortgezet onderwijs leerjaar 1); ß (de gestandaardiseerde regressiecoëfficiënt); → (het effect van M of H op een onderwijspositiekenmerk); all. (allochtone leerlingen), aut. (autochtone leerlingen).

3. BESCHRIJVINGSCATEGORIEËN

Wat *bron* betreft valt op dat lopende cohort-onderzoeken, zoals VOCL'89 en de cohorten van de Landelijke Evaluatie van het Onderwijsvoorrangsbeleid (LEO), in belangrijke mate het basismateriaal leveren voor de analyses. In feite kunnen al de beschreven onderzoeken teruggevoerd worden op elf databestanden. Duidelijk wordt ook dat de wetenschappelijke discussie over het onderhavige onderwerp van recente datum is: op vier na, zijn alle publikaties in 1990 of later verschenen. De *dataverzameling* heeft in de meeste gevallen rond 1988 plaatsgevonden of is toen van start gegaan en loopt nog steeds. In het algemeen wordt gebruik gemaakt van omvangrijke tot zeer omvangrijke *steekproeven*. In een aantal gevallen zijn vervolgens binnen deze steekproeven weer specifieke substeekproeven getrokken. Nagenoeg alle studies zijn landelijk en kunnen in meerdere of mindere mate als representatief worden beschouwd. Een derde van de analyses heeft betrekking op de situatie in het basisonderwijs, twee derde op die in het voortgezet onderwijs.

De *operationalisatie van sociaal milieu* kent een bijzonder grote mate van variëteit. Favoriete indicatoren voor milieu zijn opleiding en beroep van één of van beide ouders. Meestal betreft het vrij grove indelingen, waarbij de inschaling doorgaans wordt overgelaten, c.q. toevertrouwd aan de ouders zelf, aan de groepsleerkracht of aan de directeur van de school. Bij opleiding gaat men veelal uit van het hoogste opleidingstype dat de ouder heeft bezocht. Als alternatief kiest men ook vaak voor het hoogst voltooide type. Qua beroep vindt in enige studies een beperking plaats tot het beroep van de vader, kostwinner of van het hoofd van het huishouden. De reden dat het beroep van de moeder niet wordt meegenomen is gelegen in het feit dat een groot deel van de moeders geen betaald werk verrichten of slechts voor een beperkt aantal uren een baan hebben. Dit leidt mogelijk tot problemen bij de inschaling van het niveau. In nagenoeg alle gevallen wordt dit opgelost door moeders zonder betaald werk ('huisvrouwen') in de laagste beroepscategorie onder te brengen. Een steeds groter probleem bij de inschaling vormt het werkloos-zijn (van zowel de vaders als moeders). Ook hier is inschaling in de laagste beroepscategorie een veel voorkomende oplossing. Omdat dit echter niet altijd als even terecht wordt beschouwd, vraagt men in het geval van werkloosheid ook wel naar het laatst uitgeoefende beroep en wordt de inschaling daarop gebaseerd. In verband met deze problematiek neemt men soms een extra variabele mee, namelijk het al-dan-niet hebben van betaald werk. Hoewel het opleidings- en beroepsniveau van de vader en moeder regelmatig als afzonderlijke factoren in de analyses worden opgenomen, opteert men vaak uit efficiëntie-overwegingen voor een 'indikking'. Die wordt bereikt door het hoogste opleidings- en hoogste beroepsniveau binnen een gezin te nemen, door het opleidings- en beroepsniveau van de vader en moeder te middelen, of door factor-analyse of HOMALS te verrichten op de onderscheiden indicatoren en vervolgens op basis daarvan scores toe te kennen[6]. In enkele studies neemt men naast het beroeps- en opleidingsniveau nog andere kenmerken in een uiteindelijke milieu-index op (bv. woondichtheid, kamers p.p.).

Ook de *operationalisatie van etnische herkomst* kent een aantal varianten. Het ambigue 'herkomstland' maakt recentelijk steeds meer plaats voor het 'geboorteland' van de leerling, de

Tabel 1. De geanalyseerde studies, naar beschrijvingskenmerken.

Bron	Dataverzameling	Steekproef	Operationalisatie milieu, herkomst	Onderwijspositiekenmerk(en)	Analysetechniek/methode	Resultaten	Conclusies
Driessen (1990a) (vgl. Driessen, 1990b)	1987/88-1988/89	N=2354 ba.o8 – vo1 landelijk.	M: beroep vader en opleiding ouders. H: herkomstland vader, resp. moeder, evt. grootouders.	Toetsprestaties taal en rekenen ba.o8; doorstroomniveau vo1.	Variantie- en regressieanalyse.	Hoewel aut. binnen alle milieu-niveaus hogere prestaties en lagere doorstroomniveaus halen dan all., zijn de verschillen slechts in één cel relevant. β M → taal, rekenen en doorstroom: .37, .40 en .58; % verkl. variantie: 25, 16 en 21. β H → taal, rekenen en doorstroom: .17, .04 en -.17; % verkl. variantie: 1, <1 en 1.	M > H (3)
Driessen (1993c)	1989/90	N=19.524 vo1 landelijk.	M: hoogste beroep en hoogste opleiding binnen gezin. H: geboorteland ouders.	Onderwijspositie vo1; toetsprestaties taal, rekenen en informatieverwerking.	Regressieanalyse.	Onderzocht wordt of de effecten van M en H op de onderwijspositie en prestaties verschillen voor OVB-gebieden met veel en gebieden met weinig all. β opleiding, beroep en H: ±.30, .15 en <.05. M verklaart 10.5 – 20.9%; H voegt daar .2 – 3.9% aan toe. Voor onderwijspositie geldt dat binnen OVG's het effect van beroep en opleiding wat lager en van H wat hoger is.	M > H (4)
Driessen & Van der Werf (1992) (vgl. Driessen, 1993a; 1993b)	1989/90	N=19.524 vo1 landelijk.	M: hoogste beroep binnen gezin en opleiding ouders. H: geboorteland ouders.	Advies; toetsprestaties taal, rekenen en informatieverwerking; onderwijspositie vo1.	Variantie- en regressieanalyse.	Tussen aut. en all. bestaat m.b.t. advies alleen in de laagste milieu-categorie een relevant verschil (all. > aut.). β opleiding vader, opleiding moeder, beroep en H: .23, .21, .14 en -.04. De drie milieu-variabelen verklaren 21.8%, terwijl H daar niet meer dan .2% aan toevoegt. Bij omkering van de volgorde verklaart H 1.1% en M vervolgens nog 20.9%. Analyses m.b.t. prestaties en onderwijspositie leiden tot vergelijkbare resultaten.	M > H (5)

Bron	Dataverzameling	Steekproef	Operationalisatie milieu, herkomst	Onderwijspositiekenmerk(en)	Analysetechniek/methode	Resultaten	Conclusies
Driessen & Van der Werf (1994)	1989/90-1991/92	N=19.524 vo1-3 landelijk.	M: hoogste beroep en opleiding binnen gezin. H: geboorteland ouders.	Onderwijspositie vo3 (leerjarenladder); toetsprestaties taal en wiskunde; intelligentie.	Regressieanalyse.	β opleiding, beroep en H → onderwijspositie: .34, .15 en -.01, met 17.9, 1.4 en .0% verkl. variantie. Analyses met prestaties en IQ als afhankelijke variabelen leiden tot vergelijkbare resultaten.[1]	M > H (4)
Van Esch (1983)	1982	N=1271[2] lo6 landelijk.	M: beroepsgroep kostwinner. H: herkomstland ouders.	Deelname Cito-Eindtoets; advies; taalvaardigheid; toetsprestaties Cito-Eindtoets.	Regressieanalyse.	Binnen de totale groep verklaart H van deelname, advies en taalvaardigheid resp. 8.5, 2.5 en 4 keer zoveel variantie als M. Binnen de groep deelnemers echter verklaart H van taalvaardigheid en prestaties dubbel zo veel als M. Van advies verklaart M iets meer dan H. Het aantal jaren gevolgd Nederlands onderwijs verklaart evenveel of meer dan H en M.	H > M (3); H > M (2); M > H (1)
Van 't Hof & Dronkers (1993) (vgl. Van 't Hof & Dronkers, 1992)	1989/90	N=1898[3] vo1 landelijk.	M: beroep en opleiding ouders, betaald werk. H: geboorteland ouders en kind, etnische samenstelling gezin.	Toetsprestaties taal, rekenen en informatieverwerking; advies; onderwijspositie vo1.	(Co)variantieanalyse.	M.u.v. rekenen (β .09) is het effect van H steeds significant (β .13 – .18), maar de verkl. variantie is maximaal 3%, terwijl voor bv. taal in totaal 23% variantie wordt verklaard.	M > H (5)
Hofman (1993)	1990, 1992	N=3066 (meting 1) N=2565 (meting 2) vo2-4 landelijk (?)	M: beroep en opleiding ouders. H: geboorteland ll., ouders, grootouders en zelfidentificatie ll.	Onderwijsrendement (leerjarenladder); studieperspectief; drop-out; rekenprestaties.	Tabel-analyse; pad-analyse.	Tabel-analyse laat zien dat er geen systematische effecten zijn van M op rendement, perspectief en drop-out. Pad-analyse met prestaties, rendement en perspectief als afhankelijke variabelen bevestigt dit: het effect van M is de helft tot een derde van dat van H. Bovendien blijkt het belang van H toe te nemen naarmate het M hoger is.	H > M (4)

Bron	Dataverzameling	Steekproef	Operationalisatie milieu, herkomst	Onderwijspositiekenmerk(en)	Analysetechniek/methode	Resultaten	Conclusies
De Jong & Tacq (1985)	1981	N=408 vo, 15-jarigen Rotterdam.	M: beroep en opleiding ouders, boekenbezit, woondichtheid. H: geboorteland ouders.[4]	Bereikt onderwijsniveau op 15-jarige leeftijd (schooltype en leerjaar).	Log-lineaire analyse; logit-analyse.	M, H en verblijfsduur hebben elk een zelfstandig effect, maar het effect van M is groter dan dat van H, en dat is weer groter dan dat van verblijfsduur.	M > H (1)
De Jong (1987)	1981-1984	N=±1000 lo5 – vo2 Rotterdam.	M: beroep en opleiding ouders. H: geboorteland ouders.	IQ-test (lo6); advies; bereikt opleidingsniveau (vo2).	Variantie- en regressie-analyse; pad-analyse.	Binnen dezelfde milieu-categorie is de IQ-score van all. veel lager dan die van aut. M en H hebben elk een zelfstandig effect, maar dat van M is sterker dan dat van H.[5] *r* M * advies, resp. niveau: .40, resp. .41; *r* H * advies, resp, niveau: -.13; β M → IQ, resp. advies: .30, resp. .17; β H → IQ, resp. advies: -.20, resp. <.10.	M > H (3)
Jungbluth, Van Langen & Vierke (1990)	1988	N=1179[6] vo3 landelijk.	M: beroep en opleiding ouders. H: geboorteland ouders.	Onderwijspositie vo3 (schooltype, vakkenpakket, niveau Nederlands); advies vo.	Variantie-analyse.	Onderwijspositie kan worden verklaard door effecten van M en advies; H speelt daarbij geen rol. Als i.p.v. advies de Cito-scores worden genomen, treedt er wel een effect op van H. Dit is even groot als dat van M[7]. Ook als advies wordt voorspeld op basis van Cito-scores, M en H, dan hebben M en H vergelijkbare effecten.	M > H (1); M ≈ H (2)
Jungbluth, Verhaak & Driessen (1990)	1987/88	N=2354 ba.o8 landelijk.	M: beroep vader en opleiding ouders. H: herkomstland vader, resp. moeder, evt. grootouders.	Advies.	(Co)variantie-analyse.	Advies wordt in stap 1 voor 9.3% door H verklaard. Na invoering van taal- en rekenprestaties in stap 2 (40.1%) verdwijnt het effect van H. Als in stap 3 ook M (2.8%) wordt ingevoerd, keert het effect van H terug (1.6%). Na toevoeging van leerkrachtverwachting en onderwijsondersteunend thuisklimaat verdwijnt het effect van H weer, terwijl het effect van M hetzelfde blijft.	M > H (1)

Bron	Dataverzameling	Steekproef	Operationalisatie milieu, herkomst	Onderwijspositiekenmerk(en)	Analysetechniek/methode	Resultaten	Conclusies
Kerkhoff (1988)	1983-1984	N=179 lo5-6 Tilburg, Breda.	M: opleiding en beroep ouders. H: herkomstland ouders, grootouders.	Advies; schoolprestaties.	Tabel-analyse; variantieanalyse.	Binnen een vergelijkbaar M zijn de adviezen van all. lager dan die van aut. Na controle voor M is er alleen een significant verschil voor lezen (Turken en Marokkanen vs. rest).	M > H (2)
Kerkhoff & Dronkers (1990) (vgl. Dronkers & Kerkhoff, 1990)	1983-1984	N=179 lo5-6 Tilburg, Breda.	M: opleiding en beroep ouders. H: herkomstland ouders, grootouders.	Toetsprestaties taal en rekenen.	Regressieanalyse.	Ook bij gelijke milieu- en andere achtergrondkenmerken zouden all. een geringe achterstand blijven houden t.o.v. aut. Bij een gelijk opleidings- en beroepsniveau zou 76% van de taal-, en 15% van de rekenachterstand verdwijnen.	H > M (2)
Koeslag & Dronkers (1993) (vgl. Koeslag & Dronkers, 1994)	1989/90-1991/92	N=5057[3] vo1-3 landelijk.	M: beroep en opleiding ouders. H: geboorteland en nationaliteit ouders.	Toetsprestaties taal, rekenen en informatieverwerking; onderwijspositie in 1e, 2e en 3e schooljaar vo; overadvisering.	LISREL.	Na controle voor M (β .54) resteren er nauwelijks nog directe effecten van H ($\beta \leq$-.06) op prestaties. Ook overadvisering is vooral een milieu-effect (β M: .44, en H: ≤.19). De directe effecten van M en H op de onderwijsposities zijn alle ≤.05.	M > H (7)
Kremers (1990)	1988	N=27.093 vo3 landelijk.	M: opleiding ouders. H: spreektaal ll. met moeder.	Toetsprestaties biologie, Engels en wiskunde.	Tabel-analyse.	Voor alle drie de vakken presteren aut. uit eenzelfde M (laag vs. hoog) beter dan all. Hoewel de verschillen dus niet grotendeels door M worden bepaald, is het wel zo dat de verschillen voor het lage niveau meer dan dubbel zo groot zijn dan voor het hoge niveau. Voor Engels zijn de verschillen bij het hoge niveau nog maar minimaal.	H > M (3)

Bron	Dataverzameling	Steekproef	Operationalisatie milieu, herkomst	Onderwijspositiekenmerk(en)	Analysetechniek/methode	Resultaten	Conclusies
Van Langen & Jungbluth (1990a)	1988	N=1087[8] vo3 landelijk.	M: beroep en opleiding ouders. H: geboorteland ouders.	Onderwijspositie vo3 (schooltype, vakkenpakket en niveau Nederlands).	Variantie-analyse.	Van de onderwijspositie verklaart M het leeuwedeel van de variantie, maar ook godsdienst, verblijfsduur en spreektaal zijn van belang. Echter na controle voor M, blijft er voor de overige variabelen nagenoeg niets over. Binnen de verhouding M en H verklaart H na controle voor M nog 4% en M na controle voor H nog 10%.	M > H (1)
Van Langen & Jungbluth (1990b)	1988	N=2156 vo3 landelijk.	M: beroep en opleiding ouders. H: geboorteland ouders.	Onderwijspositie vo3 (schooltype, vakkenpakket en niveau Nederlands).	Variantie-analyse.	M verklaart 14% en H 2%, met een overlap van 3%.[9]	M > H (1)
Latuheru & Hessels (1993) (vgl. Latuheru & Hessels, 1994b)	1989	N=319[10] vo2 landelijk (?)	M: opleiding ouders. H: geboorteland ouders.	Toetsprestaties rekenen.	(Co)variantie-analyse.	Na controle voor M blijft er een zelfstandig effect van H. Na controle voor M en intelligentie is er echter geen effect meer van H. Omgekeerd is er na controle voor H en intelligentie ook geen effect van M. De verschillen in prestaties kunnen dus worden verklaard door intelligentie en de variantie die H en M gemeenschappelijk hebben. H is dus niet van invloed op de prestaties: het verband tussen H en prestaties betreft nl. een schijnverband a.g.v. verschillen in M en intelligentie. (In 1994b is deze conclusie gewijzigd: a.g.v. van de contaminatie van M en H kan het effect van H op prestaties niet eenduidig worden vastgesteld.)	M = H (1)

Bron	Dataverzameling	Steekproef	Operationalisatie milieu, herkomst	Onderwijspositiekenmerk(en)	Analysetechniek/methode	Resultaten	Conclusies
Latuheru & Hessels (1994a)	(1) 1984; (2) 1992	(1) N=174 vo2 Rotterdam; (2) N=186[11] vo2 landelijk (?).	M: opleiding ouders. H: geboorteland ouders.	(1) Bereikt opleidingsniveau; (2) toetsprestaties rekenen.	(Co)variantieanalyse.	In (1) is er na controle voor M geen effect meer van H, en na controle voor H geen effect van M. Bij simultane schatting hebben M en H beide geen effect. In (2) is er na controle voor M nog steeds een effect van H. Na controle voor H is er echter geen effect meer van M. Bij simultane schatting is er wel een effect van H, maar niet van M. Na scheiding van de unieke en gemeenschappelijke delen van de variantie zijn de analyses herhaald. Dan blijkt in (1) alleen het gemeenschappelijke deel significant. In (2) is er wel een uniek effect van H en tevens van H en M gemeenschappelijk. In zowel (1) als (2) is de door H en M samen verkl. variantie aanzienlijk groter dan de uniek verkl. variantie (5 vs. 1 en 2%, resp. 11 vs. 6 en 1%). A.g.v. multicollineariteit zijn unieke effecten van H en M dus niet-significant of relatief klein.	M ≈ H (1); H > M (1)
Mulder & Tesser (1992)	1989	N=5322 vo1 landelijk.	M: opleiding ouders en beroep vader. H: herkomstland vader.	Toetsprestaties taal en rekenen (ba.o8); advies; brugklasniveau.	Regressieanalyse.	Als na H opleiding vader wordt ingevoerd, is het effect van H op advies en brugklasniveau niet meer significant. Na toevoeging van opleiding moeder en beroep vader wordt het effect van H nog kleiner. Wanneer m.b.t. de prestaties eerst H en daarna M wordt ingevoerd, dan blijven er echter nog wel effecten van H. Voor taal is het effect van H twee keer zo groot als dat van M; voor rekenen is het ongeveer gelijk. Wanneer echter de volgorde verandert – eerst M en dan H – dan is de verhouding het omgekeerde.	M ≈ H (2); H > M (1); M ≈ H (1)

Bron	Dataverzameling	Steekproef	Operationalisatie milieu, herkomst	Onderwijspositiekenmerk(en)	Analysetechniek/methode	Resultaten	Conclusies
Pels & Veenman (1994)	1991	N=1633 12-19-jarigen 15 gemeenten.	M: ? H: ?	Scholingsniveau.	Matching[12].	De oorspronkelijke samenhangen tussen scholingsniveau en etniciteit dalen na matching op leeftijd en M (voor Turken van .29 naar .18; voor Marokkanen van .31 naar .21). Toevoeging van andere migratiekenmerken leidt bij de ene all. groep vervolgens weer tot een stijging van de samenhang en bij de andere tot een verdere daling. Hoewel M in belangrijke mate bepalend is, worden de verschillen in onderwijspositie daardoor niet volledig wegverklaard.	M > H (1)
Roelandt, Martens & Veenman (1990) (vgl. Roelandt & Veenman, 1990)	1988	N=1465 12-19-jarigen 10 gemeenten.	M: opleiding en beroep hoofd huishouden en aantal kamers p.p. H: nationaliteit hoofd huishouden, geboorteland.	Bereikt scholingsniveau.	Logistische regressie-analyse.	Verschillen in M verklaren voor een belangrijk deel, maar zeker niet volledig de scholingsachterstanden van all. Er zijn daarbij echter grote verschillen tussen en ook binnen etnische groepen. Ook wanneer na M nog wordt gecontroleerd voor andere leerling- en migratiekenmerken[13] blijft er voor all. een geringere kans om een gunstig scholingsniveau te bereiken.	M > H (1)
Schouten (1989)	1982-1988	N=306[14] lo6 + 6 jaar landelijk.	M: beroep en opleiding vader. H: nationaliteit ouders.	Bereikt onderwijsniveau na 5 jaar vo; jaar van uitstroom.	Matching[15].	All. hebben een minder succesvolle loopbaan en verlaten eerder het onderwijs dan aut., ook na controle voor beroep vader, opleiding vader, geslacht, gemeentegrootte en Cito-score. Echter, op het moment van uitstroom blijkt er geen significant verschil meer te zijn in bereikt niveau. Het lagere niveau valt dus te verklaren uit het eerder verlaten van het onderwijs.	H > M (1); M ≈ H (1)

Bron	Dataver-zameling	Steekproef	Operationalisatie milieu, herkomst	Onderwijspositiekenmerk(en)	Analysetechniek/methode	Resultaten	Conclusies
Tesser, Mulder & Van der Werf (1991)	1988/89	N=38.500 ba.o4,6,8 landelijk.	M: opleiding ouders, beroep vader. H: herkomstland vader.	Toetsprestaties taal en rekenen.	Regressie-analyse.	Na controle voor M blijft een sterk effect van H bestaan. Bij taal bindt H evenveel variantie als de drie M-indicatoren samen; bij rekenen is de bijdrage minder. β opleiding vader, opleiding moeder, beroep vader, H op taal in groep 4: .10, .11, .04, .31. M verklaart 9.3% variantie en H voegt daar 9.1% aan toe.	H ≈ M (1); M > H (1)
Tesser & Vierke (1990)	1988/89	N=3553[16] ba.o8 landelijk.	M: beroep, opleiding en betaald werk vader. H: herkomstland vader.	Toetsprestaties taal en rekenen.	Tabel-analyse; variantie-analyse.	Een vergelijking van de prestaties van vijf etnische groepen binnen de opleidings- en beroepsniveaus. Voor taal blijven verschillen bestaan van .75 à 1 *sd*; voor rekenen zijn ze .5 *sd*. Uit variantie-analyse blijkt dat het effect van H voor taal twee keer zo groot is als dat van beroep/werkloosheid. Voor rekenen is echter het effect van beroep/werkloosheid iets groter dan dat van H. Voor opleiding worden soortgelijke effecten gevonden.	H > M (1); M > H (1)
Tesser, Van der Werf, Mulder & Weide (1989)	1988	N=3623 ba.o8[17] landelijk.	M: beroep en opleiding ouders en betaald werk. H: geboorteland ouders.	Advies.	Tabel-analyse; pad-analyse.	Na controle voor opleiding vader verdwijnt het verband tussen advies en beroep en ook tussen advies en H. In een pad-analytisch model met H als exogene en opleiding en beroep als endogene variabelen gaat er alleen een significant effect uit van H op opleiding (β .33). Het effect van H op advies is nihil (β .00). Het directe effect van opleiding op advies daarentegen is .39 en het indirecte, via beroep, is .05. Het effect van beroep op advies bedraagt .08. Conclusie: achterstanden van all. in termen van advies kunnen volledig worden toegeschreven aan het opleidingsniveau van de vaders.	M > H (1)

Bron	Dataver-zameling	Steekproef	Operationalisatie milieu, herkomst	Onderwijsposi-tiekenmerk(en)	Analysetech-niek/methode	Resultaten	Conclusies
Wolbers & Driessen (1994)	1989/90-1991/92	N=2024[18] vo1-3 landelijk.	M: hoogste beroep en hoogste opleiding binnen gezin. H: geboorteland ouders.	Advies; toets-prestaties taal en rekenen vo1; onderwijspositie schooljaar 3.	LISREL.	Opleidingsniveau heeft een direct effect op advies en toetsprestaties van all., maar niet meer op onderwijspositie in jaar 3. Hoewel deze directe effecten gedurende de loopbaan afnemen, blijven de totale effecten gelijk. Van beroepsniveau gaat alleen een direct effect uit op advies en indirecte effecten op prestaties en onderwijspositie. H heeft alleen een directe invloed op de taalprestaties. De invloed van M is bij dit alles aanzienlijk groter dan die van H.	M > H (4)

Bijzonderheden:

1. Toevoeging van verblijfsduur levert geen extra verklaarde variantie op.
2. Het betreft alleen allochtone leerlingen met een redelijke beheersing van het Nederlands.
3. Dit betreft een selectie uit de oorspronkelijke steekproef van N=19.524.
4. Verblijfsduur is als aparte variabele meegenomen.
5. Ook na toevoeging van verblijfsduur en boekenbezit.
6. Dit betreft een selectie uit de oorspronkelijke steekproef van N=2156.
7. Deze analyses zijn verricht bij een deel van de steekproef.
8. Dit betreft een selectie van de allochtone leerlingen uit de oorspronkelijke steekproef van N=2156.
9. Toevoeging van onderwijsondersteunend thuisklimaat leidt niet tot verhoging van de verklaarde variantie.
10. Dit betreft een selectie uit de oorspronkelijke steekproef van N=3066, nl. 9% van de Nederlanders en alleen leerlingen die ≥25% van de IQ- en rekenitems goed hadden.
11. Dit betreft een selectie uit de oorspronkelijke steekproeven van (1) N=±1000 en (2) N=2565 (zodanig dat de aantallen per etnische subgroep gelijk zijn).
12. N=640 paren (dit varieert per analyse).
13. Leeftijd, geslacht, verblijfsduur, wijze van instroom en oriëntatie op de Nederlandse samenleving.
14. Dit betreft een selectie uit de oorspronkelijke steekproef van N=14.987.
15. N=153 paren.
16. Dit betreft een selectie uit een oorspronkelijke steekproef van N=38.500.
17. De betreffende leerlingen zaten in 1986/87 in groep 8.
18. Dit betreft een selectie van de allochtone leerlingen uit de oorspronkelijke steekproef van N=19.524.

vader, de moeder en/of grootouders. De nationaliteit als criterium wordt in verband met het toenemend aantal naturalisaties eveneens steeds minder toegepast. Maar ook het geboorteland-criterium zal niet lang meer adequaat blijven. Immers, het aantal ouders dat zèlf in het buitenland geboren is, neemt snel af. Een bijkomend probleem is de positie van kinderen uit gemengde gezinnen; onduidelijk is tot welke categorie zij gerekend moeten worden? In sommige studies wordt als oplossing gekozen voor het opnemen van de gezinssamenstelling (al-dan-niet gemengd gezin) als aparte variabele. Om genoemde redenen wordt wel voorgesteld uit te gaan van zelfidentificatie: de 'allochtoon' bepaalt zelf tot welke categorie hij/zij zich rekent. Critici wijzen in dit verband echter op het gevaar van willekeur en opportunisme. Duidelijk zal zijn dat het eenduidig definiëren van de etnische herkomst nu al een probleem vormt en dat dit met het verstrijken van de tijd alleen maar groter zal worden.

De meest geanalyseerde *onderwijspositiekenmerken* (de afhankelijke variabelen) betreffen met name: toetsprestaties in het basisonderwijs, advies voortgezet onderwijs, bereikt onderwijsniveau in het x^e jaar van het voortgezet onderwijs, toetsprestaties voortgezet onderwijs. Qua *analysetechniek/methode* ligt het accent op variantie- en regressie-analyse.

Gezien de verscheidenheid van het materiaal laten de *resultaten* zich niet gemakkelijk samenvatten. Toch valt er een algemene lijn in te ontwaren; deze ligt tussen twee polen: in sommige studies is er na controle voor milieu geen effect meer van herkomst, terwijl bij andere studies er onder constanthouding van herkomst geen invloed meer is van milieu. Het merendeel van de onderzoeksresultaten zit daar tussenin; de teneur is echter dat het milieu belangrijker is dan de herkomst. Hoewel er op de manier waarop de resultaten kwantitatief zijn samengevat in de kolom *conclusies* wellicht kritiek mogelijk is (sommige studies laten zich niet zo duidelijk samenvatten), doet dit niets af van de dominante positie van milieu. In 51 analyses (68%) legt milieu een zwaarder gewicht in de schaal dan herkomst, in 18 analyses (24%) is het omgekeerde het geval, en in 6 analyses (8%) is er geen verschil.

4. VERSCHILLEN VERKLAARD: METHODISCHE ASPECTEN

Het overzicht laat zien dat er ondanks de vele overeenkomsten in de richting van de onderzoeksbevindingen toch ook nog flink wat afwijkingen zijn. In een enkele studie wordt geen zelfstandig effect aangetroffen van milieu (bv. Hofman, 1993), in een andere juist niet van herkomst (bv. Tesser, Van der Werf, Mulder & Weide, 1989). Lang niet alle conclusies zijn even stellig; in een aantal gevallen voegen de onderzoekers relativeringen aan hun uitspraken toe. De Jong (1987) bijvoorbeeld concludeert dat milieu belangrijker is dan herkomst, maar: 'Dit neemt niet weg dat de allochtone herkomst een extra handicap oplevert bij het doorlopen van het Nederlandse onderwijs' (o.c., p. 207). En Driessen & Van der Werf (1992) stellen op basis van vergelijkbare bevindingen: 'Echter, omdat een groot deel van de allochtone leerlingen afkomstig is uit de laagste sociaal-economische milieus kan met evenveel recht worden gesproken van een sociaal als van een etnisch fenomeen. De allochtone herkomst valt hier immers samen met het lage opleidings- en beroepsniveau van de ouders' (o.c., p. 78). De eindconclusie van Van 't Hof & Dronkers (1993) luidt enerzijds dat in hoofdzaak de normale 'Nederlandse' milieueffecten verantwoordelijk zijn voor de lagere prestaties van de allochtone leerlingen, maar dat anderzijds de patronen en processen van de schoolloopbanen van allochtone en autochtone leerlingen verschillen als gevolg van de migratiegeschiedenissen en uiteenlopende culturen, en wel zodanig dat kan worden gesteld dat een aparte aanpak van allochtone leerlingen gerechtvaardigd is. Zo bezien zou dit betekenen dat '... de tegenstelling tussen klasse en etnische cultuur als verklaring van schoolloopbanen van migrantenleerlingen een valse tegenstelling is.' (o.c., p. 19). En Kerkhoff (1988) vermoedt '... dat de factor "allochtoon" het negatieve effect van een laag sociaal-economisch milieu versterkt' (o.c., p. 79).

Deze relativeringen doen weinig af aan het feit dat er grote discrepanties in bevindingen bestaan. Het is de vraag waardoor deze kunnen worden verklaard. Deze paragraaf biedt enkele –

met name methodische – aanzetten, maar heeft nadrukkelijk niet de pretentie het definitieve antwoord te geven – daarvoor is de materie – zoals hierna eens te meer zal blijken – te complex.[7]

4.1. Tijd

Een aspect dat waarschijnlijk niet van invloed is op de verschillen in resultaten betreft het moment van dataverzameling. Het merendeel van de gegevens stamt namelijk uit dezelfde, relatief korte periode van 1988 tot 1992. Een ander tijdselement dat mogelijk wèl van invloed is betreft de onderwijsfase waarop de studie betrekking heeft: gaat het om het basisonderwijs, de overgang van basis- naar voortgezet onderwijs, het voortgezet onderwijs, of een latere fase? De jongste steekproef waarbij metingen zijn verricht betreft leerlingen in groep 4 (Tesser, Mulder & Van der Werf, 1991), terwijl de oudste steekproef bestaat uit 19-jarigen (Roelandt, Martens & Veenman, 1990). Tussen deze leeftijdscategorieën in ligt een periode van ruim tien jaar. Het is daarbij zeker niet uitgesloten dat er effecten van onderwijservaring (evt. in combinatie met verblijfsduur) optreden. Anderzijds is het ook niet ondenkbaar dat deze effecten gerelateerd zijn aan specifieke kenmerken van de onderwijsfase, zoals het aanbod, de inhoud en toelatingseisen.

4.2. Milieu en herkomst

Evident is dat er in principe nogal wat variatie mogelijk is in de operationalisatie van de onafhankelijke variabelen. Met name als het gaat om de definiëring van de allochtone groep kan dit tot forse verschillen in de omvang van de verschillende categorieën leiden. Het maakt bijvoorbeeld nogal wat uit (qua aantallen, maar vervolgens bv. ook qua gemiddelde prestaties van de onderscheiden groepen) of de categorie Surinamers wordt samengesteld op basis van het nationaliteits- of op basis van het geboorteland-criterium; iets soortgelijks geldt voor inmiddels genaturaliseerde migranten. Toch lijkt het er niet op dat voor de hier bestudeerde onderzoeken de operationalisatie van herkomst op zich van invloed is. Immers, nagenoeg alle studies hanteren het geboorteland-criterium.

Met betrekking tot de indicator opleiding wijzen Meijnen & Riemersma (1992) er op dat de afwijkende resultaten van Tesser, Mulder & Van der Werf (1991) mogelijk het gevolg zijn van onvoldoende differentiatie in de laagste categorie van het opleidingsniveau van de ouders. Deze categorie zou voor allochtonen te grof zijn, omdat voor de groep de verschillen niet zozeer te maken hebben met het al-dan-niet voltooid hebben van schooltypen (bv. lbo vs. mavo), maar met het aantal voltooide leerjaren binnen schoolniveaus (bv. 2 vs. 4 jaar lager onderwijs). Nadere analyses maken duidelijk dat inderdaad binnen het laagste niveau de allochtone ouders minder onderwijs hebben gevolgd dan autochtone ouders. Toch kan dit volgens Tesser (1993) niet de verklaring zijn, want allochtone leerlingen presteren in vrijwel alle opleidingsniveaus lager dan autochtone leerlingen. Bovendien geldt ook hier weer dat de constatering dat er eigenlijk onvoldoende differentiatie is in opleidingsniveaus – hoewel op zich volkomen terecht – waarschijnlijk geen verklaring kan bieden voor de gerapporteerde onderzoeksresultaten; praktisch alle operationalisaties lijden namelijk aan dit euvel.

Wat zonder meer een probleem vormt bij het onderhavige type onderzoek is het grote aantal ontbrekende waarden (de 'missings'). Met name als het verkrijgen van informatie over herkomst en milieu afhankelijk is van de medewerking van de betreffende ouders, blijkt dat relatief veel van hen deze – meestal via schriftelijke vragenlijsten verzamelde – gegevens niet kunnen of willen verstrekken. Het is niet ongewoon dat juist op deze variabelen twintig tot veertig procent missings voorkomen (Tesser, Van der Werf, Mulder & Weide, 1989; Van 't Hof & Dronkers, 1993). In sommige gevallen zijn daarom representativiteitsanalyses uitgevoerd (Driessen & Van der Werf, 1992; Tesser, Mulder & Van der Werf, 1991). Deze wijzen er niet op dat de groep leerlingen van wie de betreffende gegevens ontbreken op de afhankelijke variabelen afwijkt van de groep van wie ze wel bekend zijn. Ondanks deze 'geruststelling' blijven er toch twijfels over het effect van deze grote aantallen missings. De verwachting is namelijk dat het een specifieke groep betreft; er zal waarschijnlijk een oververtegenwoordiging zijn van allochtone ouders die de Nederlandse taal niet of onvoldoende beheersen en om die reden de vragenlijst niet kunnen invullen. Daar staat echter tegenover dat er aanwijzingen zijn dat tegelijkertijd

juist hoogopgeleide ouders weigeren dergelijke vragenlijsten in te vullen vanwege de privacygevoeligheid van de informatie. Verder moet worden opgemerkt dat dit probleem van de missings in alle onderzoeken speelt; in dat opzicht kan het dus moeilijk een verklaring bieden voor de aangetroffen onderzoeksbevindingen.

4.3. Onderwijspositie-kenmerken

In de onderzoeken zijn verschillende onderwijspositie-kenmerken geanalyseerd. De vaststelling van het advies en het bereikte onderwijsniveau in het voortgezet onderwijs heeft meestal op analoge wijze plaats gevonden. Dat wil zeggen dat men de schooltypes ordent van laag naar hoog en er vervolgens een rangordescore aan toekent. Deze wordt voor de variantie- en regressie-analyses opgevat als een variabele op intervalniveau (bv. De Jong, 1987). In een enkel geval kiest men voor een iets gedifferentieerdere aanpak, namelijk door behalve het schooltype ook het leerjaar (de leerjarenladder) en bijvoorbeeld het vakkenpakket en taalniveau te verdisconteren (bv. Driessen & Van der Werf, 1994, resp. Jungbluth, Van Langen & Vierke, 1990). Aangezien deze procedures veel gelijkenis vertonen, valt echter niet direct te verwachten dat ze de verschillen in analyse-resultaten kunnen verklaren.

Voor de bepaling van het prestatieniveau is in het algemeen gebruik gemaakt van toetsen, meestal voor de basisvaardigheden taal en rekenen. Van belang daarbij zijn aspecten als inhoud, bereik en eventuele bias. Zo zijn bijvoorbeeld de toetsen die worden gehanteerd bij de evaluatie van het OVB qua moeilijkheidsgraad speciaal afgestemd op de doelgroepen van het OVB (bv. Mulder & Tesser, 1992). Dit betekent dat deze toetsen vooral differentiëren aan de onderkant en wat minder gericht zijn op de betere leerlingen; voor deze laatste groep treden daardoor wellicht plafondeffecten op. Vooralsnog is onduidelijk wat daar de consequenties van zijn. Op een paar uitzonderingen na (bv. Driessen, 1990a; Latuheru & Hessels, 1993), worden in de meeste studies qua inhoud 'standaard'-toetsen afgenomen (bv. Van Esch, 1983; Kremers, 1990). Er is daarbij dus geen rekening gehouden met specifieke problemen die allochtone leerlingen mogelijk hebben met het maken van toetsen. Dergelijke problemen kunnen op verschillende vlakken liggen. Allochtonen zijn mogelijk in het nadeel doordat zij niet of minder (dan autochtonen) bekend zijn met het maken van toetsen sec (de toetscultuur). Voor leerlingen die vanaf het kleuteronderwijs aan het Nederlandse onderwijs hebben deelgenomen (en dat is een groeiende meerderheid) ligt dit echter niet voor de hand, misschien voor zij-instromers wat meer. Wat eventueel wèl van invloed kan zijn is toets- en itembias, die bijvoorbeeld kan ontstaan als gevolg van het nog niet voldoende beheersen van de Nederlandse taal en door voor de allochtone leerlingen onbekende, typisch Nederlandse culturele item-inhouden. Of dit inderdaad het geval is, valt vooralsnog moeilijk aan te tonen. Uiterwijk (1994) concludeert dat onderzoek naar dit fenomeen met veel onzekerheden is omgeven; eenduidige resultaten heeft het dan ook nog nauwelijks opgeleverd. Het voorafgaande samenvattend zijn er goede gronden om te veronderstellen dat bepaalde toetsaspecten van invloed kunnen zijn op de vergelijking van de prestaties van allochtone en autochtone leerlingen.

Misschien niet helemaal aansluitend bij het voorafgaande, maar in ieder geval wel interessant met betrekking tot het geconstateerde differentiële belang van milieu en herkomst in relatie tot de verschillende onderwijspositie-kenmerken (bv. Mulder & Tesser, 1992) zijn de opmerkingen van Van 't Hof & Dronkers (1993). Op basis van hun analyses merken zij op dat de etnische factor vooral afwezig is bij afgeleide zaken als advies en onderwijspositie, waarbij niet alleen de resultaten van de leerlingen zelf, maar ook de keuzen van onderwijsgevenden en ouders een rol spelen. 'Verschillen in schoolloopbanen zouden zo meer een gevolg zijn van verschillen in houding van volwassenen ten opzichte van het autochtone respectievelijk allochtone kind, dan van verschillen tussen de leerlingen' (o.c., p. 18). Daar kan natuurlijk aan worden toegevoegd dat het etnische effect zich vooral zal openbaren bij die onderwijsaspecten, die welhaast per definitie problematisch zijn voor allochtonen. Aangezien de beheersing van de Nederlandse taal de meeste moeilijkheden oplevert, zullen zich daar dan ook sterkere effecten van etnische herkomst voordoen, dan bij rekenen (bv. Mulder & Tesser, 1992; Tesser & Vierke, 1990).

4.4. Analyse

In de meeste onderzoeken wordt gebruik gemaakt van variantie- of regressie-analyse. Deze technieken stellen specifieke eisen aan bijvoorbeeld het meetniveau en de verdeling van de variabelen. Met name wat betreft de operationalisatie van herkomst doen zich meerdere problemen voor. Op de eerste plaats is het zo dat min of meer representatieve steekproeven onherroepelijk geconfronteerd worden met zeer ongelijke aantallen per categorie. De empirie is nu eenmaal dat slechts rond de tien procent van de leerlingen van allochtone herkomst is, en dat binnen deze groep vaak nog verschillende subgroepen (zouden moeten) worden onderscheiden. Ofschoon de totale steekproef meestal vrij omvangrijk is, betekent dit toch dat onderscheiden subgroepen soms niet meer dan enkele tientallen allochtone leerlingen omvatten (tegenover vele duizenden autochtone leerlingen). Bij zeer ongelijke celvullingen wordt binnen variantie-analyse minder snel voldaan aan de assumptie van homogene varianties, hetgeen kan leiden tot het ten onrechte verwerpen van de nul-hypothese. Bovendien is er bij ongelijke aantallen in de cellen sprake van een non-orthogonaal design. Als gevolg hiervan bestaat er overlap in variantie; dezelfde variantie kan dus worden toegeschreven aan meer dan een kenmerk. Tegelijkertijd speelt het probleem van de geringe absolute aantallen allochtonen. Dit laatste kan leiden tot onstabiele, onbetrouwbare analyse-resultaten (Tabchnick & Fidell, 1989). In sommige onderzoeken pakt men deze problematiek aan door middel van oversampling: men zorgt er voor dat er 'voldoende' allochtone leerlingen in de steekproef vertegenwoordigd zijn (bv. Tesser, Van der Werf, Mulder & Weide, 1989[8]). Een andere aanpak is die waarbij categorieën van allochtonen op grond van kwantitatieve en kwalitatieve overwegingen worden samengevoegd, hetgeen in het verst doorgevoerde geval leidt tot de dichotomie allochtoon versus autochtoon (bv. Driessen, 1990a). Hoewel dit misschien in de richting van een oplossing gaat, zal duidelijk zijn dat de ongelijke celvulling in feite blijft bestaan. Bovendien is het de vraag in hoeverre het samenvoegen recht doet aan de geconstateerde onderlinge verschillen tussen de diverse allochtone groepen (Van 't Hof & Dronkers, 1993). Voor regressie-analyse komt daar waarschijnlijk nog een probleem bij. Als in een dergelijke analyse milieu is vertegenwoordigd met bijvoorbeeld zeven categorieën en herkomst (na indikking) met slechts vier of zelfs maar twee, dan zal dit – louter als gevolg van deze verschillen – de regressiecoëfficiënten en de percentages verklaarde variantie beïnvloeden. In het algemeen kan verwacht worden dat dit qua verklaringskracht in het nadeel werkt van de variabele met de minste categorieën, hier dus etnische herkomst.

Dat de bovengenoemde problemen inderdaad verantwoordelijk kunnen zijn voor een verschuiving in het relatieve gewicht van milieu en herkomst vindt bevestiging in een van de tijdens de ORD'94 gepresenteerde papers (Latuheru & De Wit, 1994). Na correctie via sampling vinden de betreffende onderzoekers inderdaad dat het belang van herkomst is toegenomen ten nadele van dat van milieu. Daar kan echter tegenover worden gesteld dat in analyses waarbij – mede om genoemde redenen – een beperking heeft plaatsgevonden tot de allochtone leerlingen, het milieu (i.c. ouderlijk opleidingsniveau) toch nog een sterker effect heeft dan herkomst (Wolbers & Driessen, 1994).

In de bestudeerde onderzoeken wordt in regressie-analyse de variabele herkomst op twee manieren behandeld. In het eerste geval voert men hem in als een serie dummies (het aantal onderscheiden herkomst-categorieën minus één – de referentiecategorie). In het tweede geval ordent men de categorieën op basis van de afhankelijke variabele, bijvoorbeeld de toetsprestaties of de leerjarenladderpositie, en kent er een rangordescore aan toe. In de analyses wordt herkomst vervolgens opgevat als een interval-niveau variabele. Hoewel onduidelijk is wat nu precies het gevolg is van de tweede benadering (in hoeverre is er sprake van contaminatie van de afhankelijke variabele en de onafhankelijke variabele 'herkomst'?), blijkt uit vergelijkende analyses dat beide typen van analyses tot identieke resultaten leiden (bv. Driessen & Van der Werf, 1992).

Een apart probleem vormt het zeer sterk gecorreleerd-zijn van de onafhankelijke variabelen, i.c. milieu en herkomst. Dit verschijnsel, dat te boek staat als multicollineariteit, kan tot verschillende problemen leiden. Omdat de onafhankelijke variabelen een groot deel van de variantie gemeenschappelijk hebben, ontstaat er een interpretatieprobleem. Hoe sterker de samen-

hang, des te minder er overblijft voor eventuele unieke bijdragen van de afzonderlijke kenmerken. Verder kan multicollineariteit er toe leiden dat de schattingen van regressie-coëfficiënten, als gevolg van de grote standaardfouten, onnauwkeurig worden (Cohen & Cohen, 1983; Tacq, 1991; Veenman, 1993). Inderdaad is er blijkens de verschillende onderzoeksverslagen sprake van hoge samenhangen. Zo bedraagt de nominaal-metrische samenhang (Eta) tussen herkomst en milieu (i.c. opleidingsniveau) in het onderzoek van De Jong (1987) .53, in dat van Hofman (.75) en in dat van Driessen & Van der Werf (1992) .50. Opmerkelijk is het wel erg grote verschil tussen de samenhang in Hofmans bestand versus die in beide andere bestanden. Waar dit aan ligt is vooralsnog onduidelijk, maar waarschijnlijk is het niet geheel toevallig dat juist in zijn onderzoek het effect van herkomst aanzienlijk sterker is dan dat van milieu, terwijl in de andere twee onderzoeken juist milieu als sterk dominant uit de bus komt. Hoewel de samenhang in Hofmans onderzoek als zeer sterk gekwalificeerd kan worden, is het de vraag in hoeverre dat doorwerkt in de resultaten. Welke norm ten aanzien van multicollineariteit gehanteerd moet worden is niet helder; sommige statistische handboeken noemen samenhangen vanaf .60, andere van .80 (vgl. Dronkers & Van 't Hof, 1993). Uiteraard dient bij de interpretatie van deze coëfficiënten ook de steekproefomvang in overweging te worden genomen. Hoewel bijvoorbeeld Tabachnick & Fidell (1989) het niet aanraden twee variabelen met een bivariate samenhang van .70 of hoger in een-en-dezelfde analyse op te nemen, treden volgens hen statistische problemen (en dan m.n. onstabiele oplossingen) echter pas op bij aanzienlijk hogere correlaties. Volgens deze methodologen zijn de hierboven gerapporteerde empirische correlatiecoëfficiënten dus geenszins problematisch.

Naar aanleiding van deze problematiek hebben Latuheru & Hessels (1994) vergelijkende analyses uitgevoerd op de data van De Jong en Hofman, waarbij via 'forced entry' de volgorde van milieu en herkomst is gemanipuleerd. Gegeven de multicollineariteit bleek dat de volgorde van invoer zeer bepalend was voor de hoeveelheid verklaarde variantie, met name voor de gegevens van Hofman. Na controle voor de wederzijdse invloed echter, kwam naar voren dat de unieke effecten van beide onafhankelijke variabelen niet significant dan wel relatief klein waren; de meeste variantie hadden ze gemeenschappelijk. Het eerste deel van deze bevindingen komt evenwel niet overeen met die van Driessen & Van der Werf (1992). Ook zij hanteerden deze procedure en vonden dat na controle voor milieu de herkomst nog nauwelijks extra variantie kon verklaren (18.0, resp. .4% extra). Hoewel er bij een omgekeerde invoer-volgorde een lichte verschuiving was opgetreden, bleef het effect van milieu veel sterker dan dat van herkomst (14.7, resp. 3.7% extra).

5. ENKELE INHOUDELIJKE OVERWEGINGEN

Via stratificatie-onderzoek (meer precies: statusverwervingsonderzoek) wordt onderzocht in hoeverre de positie van ouders van invloed is op die van hun kinderen, en daarmee op de ongelijkheid in onderwijs- en levenskansen. Als uit onderzoek naar onderwijskansen blijkt dat bij gelijke intelligentie kinderen uit het ene milieu meer kansen hebben dan kinderen uit een ander milieu, dan moet de conclusie luiden dat bepaalde milieu-categorieën kennelijk over 'middelen' beschikken waardoor het de betreffende kinderen lukt een relatief hoge onderwijspositie te bereiken (Blees-Booij, 1994). In dit type onderzoek (waarin de etnische herkomst overigens nog nauwelijks een rol van betekenis speelt – althans niet voor de Nederlandse situatie) wordt – in navolging van Bourdieu – veelal een onderscheid gemaakt naar economische en culturele aspecten (vgl. Bakker, 1992; De Graaf, 1987). Bij het eerste aspect kan gedacht worden aan het inkomen, bij het tweede aan het opleidingsniveau en kennis van de cultuur. Opgemerkt moet worden dat deze indeling niet uniform en exclusief wordt gehanteerd en ingevuld, hetgeen de discussie erover bemoeilijkt. Een andere invalshoek is die waarbij de aandacht uitgaat naar houdingen, waarden en normen met betrekking tot opvoeding, onderwijs en arbeid (Blees-Booij, 1994). Hieraan zou ook concreet opvoedings- en socialiseringsgedrag kunnen worden toegevoegd. Een vraag die in dit verband aan de orde is luidt op welke wijze

kinderen in het thuismilieu worden voorbereid op het functioneren in de hen omringende maatschappelijk werkelijkheid. En vervolgens: in hoeverre wijkt deze voorbereiding thuis af van de verwachtingen van en eisen in de maatschappij?

Welke theoretische invalshoek ook wordt gekozen, uiteindelijk gaat het er om waar 'milieu' nu eigenlijk voor staat. Welke normen en waarden, welke kennis, welk gedrag, welke middelen worden verondersteld het containerbegrip 'milieu' te vertegenwoordigen? Zoals eerder al gesteld, is de etnische factor in dit type onderzoek nog relatief weinig doorgedrongen. En is het met betrekking tot de groep autochtonen – ondanks het vele onderzoek op dit gebied – al moeilijk om aan te geven wat milieu nu precies indiceert, ten aanzien van de allochtonen is dat een nog groter probleem. Het zou bijvoorbeeld interessant zijn na te gaan wat nu precies de betekenis is, in termen van overeenkomsten en verschillen, van sociaal milieu, i.c. opleiding en beroep, voor de verschillende etnische groepen, i.c. allochtonen[9] versus autochtonen. Hoewel dit niet de inzet van het onderhavige artikel is, kunnen er naar aanleiding van de beschreven onderzoeken toch enkele opmerkingen worden gemaakt.

Zo verwacht Tesser (1993) in verband met de operationalisatie van het ouderlijke opleidingsniveau dat de inhoudelijke betekenis van de categorie 'basisonderwijs niet voltooid' bij autochtone ouders een andere is dan die bij allochtone ouders. Bij de eerste groep zou het mogelijk een indicator zijn van ontwikkelingsachterstanden of mentale gebreken; het ontbreekt deze ouders met andere woorden aan capaciteiten. Bij de tweede groep ligt het eerder voor de hand dat de ouders niet in de gelegenheid zijn geweest om onderwijs te volgen, bijvoorbeeld omdat er in hun dorp geen school was of omdat het voor vrouwen niet gewoon of nodig bevonden werd onderwijs te volgen. Veenman (1993) merkt op dat het niet ondenkbaar is dat er juist daarom onder de allochtonen veel verborgen talent aanwezig is. Onbekend is hoe dit zich zou hebben ontwikkeld als de betreffende ouders wèl in de gelegenheid waren geweest onderwijs te volgen, en evenmin is duidelijk welke invloed dit vervolgens zou hebben gehad op de schoolloopbaan van hun kinderen. Veenman (1993) en Pels & Veenman (1994) werken dit type theoretische kwesties nader uit, niet alleen voor de indicator opleiding op zich, maar ook voor etnische herkomst. Hun relaas maakt goed duidelijk dat, als het gaat om verklaren van verschillen in schoolloopbanen, de nodige voorzichtigheid geboden is bij een rechttoe rechtaan interpretatie: ogenschijnlijk dezelfde indicatoren hebben voor onderscheiden groepen verschillende betekenissen (vgl. ook Tesser, 1989). Overigens geldt dit niet alleen voor een vergelijking van allochtonen met autochtonen in Nederland. Driessen (1990b) merkt op dat de status van migranten in Nederland sterk kan verschillen van die welke ze eerst hadden in hun herkomstland. Van het platteland afkomstige Marokkanen en Turken met enkele jaren middelbare school behoorden in hun herkomstland wellicht tot de maatschappelijke elite, terwijl ze in Nederland in de laagste milieucategorieën worden ingeschaald. Meijnen & Riemersma (1992) wijzen er in zijn algemeenheid op dat de in eerste instantie voor autochtonen ontworpen, maar nu ook voor allochtonen gehanteerde, indelingen voor zowel beroep als opleidingsniveau mogelijk niet valide zijn. Ze differentiëren wellicht onvoldoende aan de onderkant, waardoor niet alleen nuanceringen verloren gaan, maar vooral een inhoudelijke vergelijking met de positie van autochtonen mank gaat.

De hierboven gesignaleerde onduidelijkheden zijn in feite maar enkele voorbeelden van mogelijke verschillen in interpretatie van het begrip milieu. Daar zouden er nog talloze aan kunnen worden toegevoegd met betrekking tot onderwerpen als socialisering, aspiraties, onderwijsondersteuning, levensstijlen, financiële hulpmiddelen, taalgebruik, communicatie, cultuur, religie en omgangsvormen.

6. TOT SLOT

In het voorgaande is een inventarisatie gepresenteerd van empirisch onderzoek waarin de verhouding sociaal milieu en etnische herkomst in relatie tot onderwijsachterstanden het onderwerp van analyse vormt. De resultaten laten zien dat deze verhouding van studie tot studie sterk kan verschillen. Toch wordt duidelijk dat milieu wezenlijk meer gewicht in de schaal legt dan

herkomst. In ongeveer tweederde van de analyses is het effect van milieu groter dan dat van herkomst. En binnen de groep indicatoren van milieu neemt de verklaringskracht van met name het ouderlijke opleidingsniveau een centrale positie in. Dit opleidingsniveau differentieert niet alleen in sterke mate tussen de autochtonen en allochtonen, maar – zo is duidelijk geworden – ook binnen de groep allochtonen.

Hoewel er veel overeenkomsten zijn aangetroffen in bevindingen, is tegelijkertijd duidelijk geworden dat consensus ontbreekt. Daarom zijn vervolgens pogingen ondernomen verklaringen te vinden voor de discrepanties. Speciaal is daarbij de aandacht gericht op methodische aspecten als tijd, steekproef, operationalisatie, onderwijspositiekenmerken en analysemethode. Evident wordt dat bij de meeste van deze factoren kritische opmerkingen geplaatst kunnen worden. Aan de verklaring van de verschillen zelf draagt dit echter nauwelijks bij. Dit komt doordat die opmerkingen, bijvoorbeeld ten aanzien van onderzoeksmatige tekortkomingen, op nagenoeg alle besproken studies van toepassing zijn. Dit neemt niet weg dat er belangrijke problemen gesignaleerd kunnen worden. Zo zijn er problemen met betrekking tot de operationalisatie en interpretatie van de onafhankelijke variabelen, vormt het soms bijzonder grote aantal missings een bedreiging voor de betrouwbaarheid, en is de constructie van adequate toetsen een lastige opgave. Ook problematisch is de soms geringe omvang van allochtone subgroepen, de ongelijke verdeling van de factor etnische herkomst en de onderlinge sterke samenhang van herkomst en milieu. Verder kan worden opgemerkt dat misschien afzonderlijke factoren niet van invloed zijn op de verschillen in bevindingen, maar dat dat in onderlinge samenhang wel het geval is.

De vraag is vervolgens: hoe nu verder te gaan? Aanvullend op de constatering van Latuheru & Hessels (1994a) dat er sprake is van een statistisch dilemma en de conclusie van Tesser (1993) dat er '... in weerwil van het grote aantal onderzoeken, op dit moment geen stellige conclusies [zijn] te trekken over de invloed van de afzonderlijke factoren als de opleiding van de ouders of de etnische achtergrond.' (o.c., p. 164) zou het tot verder onderzoek uitnodigende devies van Dronkers & Van 't Hof (1993) gehanteerd kunnen worden: 'Voorzichtig, maar niet bang' (o.c., p. 31). In dit verband liggen er meerdere mogelijkheden tot verdiepende analyses op bestaande databestanden, waarbij bijvoorbeeld de variabelen op telkens andere manieren worden geoperationaliseerd of met de omvang van de steekproeven wordt gemanipuleerd. Via dit type simulatie-onderzoek kunnen dan de consequenties van de diverse aanpakken worden doorgerekend en vergeleken. De hier gepresenteerde onderzoeken bieden ook mogelijkheden tot het verrichten van een statistische meta-analyse. Via een dergelijke analyse kan wellicht worden gekomen tot geïntegreerde statistisch onderbouwde conclusies.[10] Verder zou het, met het oog op het vinden van oplossingen voor de onderwijsachterstanden van allochtonen, zonder meer aanbeveling verdienen de relatie tussen kwantitatief en kwalitatief onderzoek te verbeteren en de verhouding milieu en herkomst in inhoudelijk opzicht uit te diepen.

NOTEN

1. Fase (1994) onderscheidt in zijn proefschrift naast deze twee nog een derde type factoren, namelijk factoren die rechtstreeks het gevolg zijn van migratie en die dus in principe tijdelijk van aard zijn (vgl. ook Ledoux, 1994). Naar aanleiding van deze indeling merkte R. Penninx tijdens de verdediging van het proefschrift op dat, als het gaat om het karakteriseren van het unieke van elk van de factoren, dit tot problemen leidt. Kenmerken die door Fase als typisch voor de etnische factor worden beschouwd, zijn dat namelijk evenzeer voor de milieu-factor.
2. Voor de volledigheid: ook bij de vaststelling van de leerlinggewichten ten tijde van de start van het OVB (vgl. Mens, 1984; Mens & Driessen, 1985) is er een discussie gevoerd over de verhouding van de verschillende milieu- en herkomstindicatoren. Een belangrijk verschil met de huidige situatie is echter dat er destijds wat de herkomst-component betreft geen empirische gegevens voorhanden waren op basis waarvan een keuze kon worden onderbouwd. Anders dan toen wordt de huidige discussie ook voornamelijk gevoerd in vakwetenschappelijke publikaties.
3. In hoeverre dit uiteindelijk tot een wijziging van de omvang van de faciliteiten voor de huidige allochtone doelgroep zal leiden, is – gezien de combinatie van allochtoon en laagopgeleid – overigens de

vraag.

4. Deze afgrenzing houdt dus in dat er weinig aandacht zal zijn voor een inhoudelijke, theoretische interpretatie van de verschillende onderzoeksresultaten.
5. Helaas draagt deze beoordeling in een enkel geval een face value-karakter. Dit heeft enerzijds te maken met het feit dat sommige studies minder expliciet gericht zijn op de vraagstelling van het onderhavige artikel; anderzijds ontbreken soms de benodigde coëfficiënten.
6. Een andere mogelijkheid, het in LISREL opnemen van een latente milieu-factor met als indicatoren opleiding en beroep (vgl. Dronkers, 1990), komt in de bestudeerde analyses niet voor.
7. Daarbij dient – wellicht ten overvloede – bedacht te worden dat de complexiteit nog aanzienlijk zou toenemen als niet alleen de kwantitatieve (waar hier dus de nadruk op ligt), maar ook meer kwalitatieve en theoretisch-filosofische argumenten in de discussie zouden worden betrokken (vgl. bv. Ledoux, 1994; Pels, 1993; Procee, 1993; Roosens, 1991).
8. In de LEO-onderzoeken wordt met twee verschillende steekproeven gewerkt: een representatieve steekproef (de zgn. referentie-steekproef) en een steekproef waarin de allochtone leerlingen oververtegenwoordigd zijn. De analyses worden verricht op beide steekproeven, al dan niet gecombineerd.
9. Bij een dergelijke vergelijking zal overigens niet volstaan kunnen worden met een tweedeling allochtoon – autochtoon. Er zal ook gedifferentieerd moeten worden tussen de verschillende allochtone (en wellicht ook autochtone) groepen (vgl. Van 't Hof & Dronkers, 1993).
10. Overigens dienen de complicaties die zich bij dergelijke analyses zullen voordoen niet onderschat, en de mogelijkheden niet overschat te worden – zeker gezien de hierboven gesignaleerde methodologische problemen (vgl. Glass, McGaw & Smith, 1981; Petersen, 1992).

LITERATUUR

Bakker, B. (1992). *Sociale afstand, sociale status. De sociale status van werkenden, niet-werkenden en vrouwen.* Nijmegen: ITS.

Blees-Booij, A. (1994). *Culturele en economische beroepsstatus van mannen en vrouwen. Een tweedimensionele ordening.* Amsterdam: Amsterdam University Press.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale NJ/London: Lawrence Erlbaum.

Commissie Allochtone Leerlingen in het Onderwijs (1992). *Ceders in de tuin. Naar een nieuwe opzet van het onderwijsbeleid voor allochtone leerlingen.* 's-Gravenhage: DOP.

Driessen, G. (1990a). Sociaal milieu en etnische herkomst als verklaring voor verschillen in onderwijspositie. *Pedagogisch Tijdschrift, 15,* (6), 349-357.

Driessen, G. (1990b). *De onderwijspositie van allochtone leerlingen. De rol van sociaal-economische en etnisch-culturele factoren, met speciale aandacht voor het Onderwijs in Eigen Taal en Cultuur.* Nijmegen: ITS.

Driessen, G. (1993a). Socio-economic or ethnic determinants of educational opportunities? Results from the education priority policy programme in the Netherlands. *Studies in Educational Evaluation, 19,* (3), 265-280.

Driessen, G. (1993b). Berusting of uitdaging? De startpositie van allochtone jongeren in het voortgezet onderwijs. *Jeugd en samenleving, 23,* (4), 163-176, 223.

Driessen, G. (1993c). Binnen en buiten onderwijsvoorrangsgebieden. Sociaal-etnische achtergronden en onderwijspositie van leerlingen in het voortgezet onderwijs. *Pedagogische Studiën, 70,* (4), 252-262.

Driessen, G. & Werf, G. van der (1992). *Het functioneren van het voortgezet onderwijs. De positie van leerlingen in het eerste leerjaar.* Nijmegen/Groningen: ITS/RION.

Driessen, G. & Werf, G. van der (1994). *Het functioneren van het voortgezet onderwijs. De schoolloopbanen van de leerlingen: stand na drie jaar.* Nijmegen/Groningen: ITS/GION.

Dronkers, J. (1990). De ontwikkelingen in het schoolloopbaanonderzoek: een terugblik op een decennium. *Tijdschrift voor Onderwijsresearch, 15,* (1), 8-22.

Dronkers, J. (1993). Culturele verscheidenheid en sociale klassen. Een reactie op 'Het belang van een cultuurbeleid in het onderwijs' van Trees Pels in Comenius 49. *Comenius, 13,* (3), 342-346.

Dronkers, J. & Hof, L. van 't (1993). Voorzichtig, maar niet bang. Een reactie op het commentaar van Justus Veenman. *Migrantenstudies, 9,* (1), 31.

Dronkers, J. & Kerkhoff, A. (1990). Sociaal milieu, taalvaardigheid en schoolsucces bij allochtonen, dialect- en standaardtaalsprekers. *Sociologische Gids, 37,* (5), 304-319.

Esch, W. van (1983). *Toetsprestaties en doorstroomadviezen van allochtone leerlingen in de zesde klas van lagere scholen.* Nijmegen: ITS.

Fase, W. (1994). *Ethnic divisions in Western European education.* Münster/New York: Waxmann.

Glass, G., McGaw, B., & Smith, M. (1981). *Meta-analysis in social research.* Beverly Hills/London: Sage.

Graaf, P. de (1987). *De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen.* Nijmegen: ITS.

Hof, L. van 't & Dronkers, J. (1992). *Onderwijsachterstanden van allochtonen: klasse, gezin of cultuur?* Amsterdam: SCO.

Hof, L. van 't, & Dronkers, J. (1993). Onderwijsachterstanden van allochtonen: klasse, gezin of etnische cultuur? *Migrantenstudies, 9,* (1), 2-25.

Hofman, W. (1993). *Effectief onderwijs aan allochtone leerlingen. Een empirische studie naar de invloed van school- en klasfactoren op de loopbaan van allochtone en autochtone leerlingen in het voortgezet onderwijs.* Delft: Eburon.

Jong, M.-J. de (1987). *Herkomst, kennis en kansen. Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs.* Lisse: Swets & Zeitlinger.

Jong, M.-J. de & Tacq, J. (1985). Het onderwijsniveau van allochtone leerlingen. *Tijdschrift voor Onderwijsresearch, 10,* (3), 130-140.

Jungbluth, P., Langen, A. van, & Vierke, H. (1990). De trefzekerheid van het advies voortgezet onderwijs bij migranten. *Tijdschrift voor Onderwijswetenschappen, 21,* (2), 90-101.

Jungbluth, P., Verhaak, C., & Driessen, G. (1990). Vervolgadviezen in relatie tot etniciteit; hoe verhouden prestaties, gezinsachtergrond en leerkrachtoordelen zich tot elkaar? *Pedagogische Studiën, 67,* (5), 231-237.

Kerkhoff, A. (1988). *Taalvaardigheid en schoolsucces. De relatie tussen taalvaardigheid Nederlands en schoolsucces van allochtone en autochtone leerlingen aan het einde van de basisschool.* Amsterdam/Lisse: Swets & Zeitlinger.

Kerkhoff, A., & Dronkers, J. (1990). Sociaal milieu, leeftijd en sekse en de taalvaardigheid in het Nederlands van allochtone en autochtone basisschoolleerlingen. *Toegepaste Taalwetenschap in Artikelen, 37,* (2), 103-119.

Koeslag, M. & Dronkers, J. (1993). *Overadvisering en de schoolloopbanen van migrantenleerlingen en autochtone leerlingen in het voortgezet onderwijs.* Paper ORD'93. Amsterdam: SCO.

Koeslag, M. & Dronkers, J. (1994). Overadvisering en de schoolloopbanen van migrantenleerlingen en autochtone leerlingen in het voortgezet onderwijs. *Tijdschrift voor Onderwijsresearch, 19,* (3), 240-258.

Kremers, E. (red.) (1990). *Overzicht van leerresultaten aan het einde van de eerste fase voortgezet onderwijs.* Arnhem: CITO.

Langen, A. van & Jungbluth, P. (1990a). *Onderwijskansen van migranten. De rol van sociaal-economische en culturele factoren.* Amsterdam/Lisse: Swets & Zeitlinger.

Langen, A. van & Jungbluth, P. (1990b). Onderwijsondersteunend thuisklimaat bij migranten. In C. Klaassen & P. Jungbluth (red.). *Onderwijs Research Dagen 1990. Onderwijs en samenleving* (pp. 293-304). Nijmegen: ITS.

Latuheru, E. & Hessels, M. (1993). *Schoolprestaties van allochtone leerlingen: de invloed van etnische herkomst.* Paper ORD'93. Rotterdam: EUR.

Latuheru, E., & Hessels, M. (1994a). *Onderwijsprestaties en de invloed van etnische en sociaal-economische herkomst: een statistisch dilemma.* Paper Zesde Sociaal-Wetenschappelijke Studiedagen 1994. Rotterdam: EUR.

Latuheru, E. & Hessels, M. (1994b). Schoolprestaties van allochtone leerlingen: de invloed van etnische herkomst. *Tijdschrift voor Onderwijsresearch, 19,* (3), 227-239.

Latuheru, E. & Wit, W. de (1994). *Wijzigingen in de relatieve invloed van de etnische en de sociaal-economische herkomst van leerlingen op hun onderwijspositie tijdens de schoolloopbaan.* Paper ORD'94. Rotterdam/Nijmegen: RISBO/ITS.

Ledoux, G., m.m.v. Deckers, P., & Koopman, P. (1994). *De inhoud van 'milieu-verschillen' bij Turkse, Marokkaanse en Nederlandse sociale stijgers.* Paper Zesde Sociaal-Wetenschappelijke Studiedagen 1994. Amsterdam: SCO.

Matthijssen, M. & Teunissen, J. (1994). *Methodische en theoretische stagnatie in onderwijssociologisch onderzoek onder allochtone leerlingen.* Paper Zesde Sociaal-Wetenschappelijke Studiedagen 1994. Utrecht: RUU.

Meijnen, G. & Riemersma, F. (1992). *Schoolcarrières: een klassenkwestie? De schoolloopbanen van leerlingen, bezien in relatie tot de invloeden die maatschappelijke en binnenschoolse determinanten hierop hebben. Een literatuurstudie.* Amsterdam/Lisse: Swets & Zeitlinger.

Mens, A. (1984). *Het onderwijsvoorrangscriterium in het basisonderwijs.* Nijmegen: ITS.
Mens, A. & Driessen, G. (1985). *Het onderwijsvoorrangscriterium in het voortgezet onderwijs.* Nijmegen: ITS.
MO&W (1994). *Uitwerkingsnotitie achterstand (concept).* Zoetermeer: Ministerie van Onderwijs en Wetenschappen.
Mulder, L. & Tesser, P. (1992). *De schoolkeuzen van allochtone leerlingen.* Nijmegen: ITS.
Pels, T. (1993). Het belang van een cultuurbeleid in het onderwijs. *Comenius, 13,* (1), 42-55.
Pels, T., & Veenman, J. (1994). *Onderwijsachterstanden bij allochtone kinderen: het ontbrekende onderzoek.* Paper Zesde Sociaal-Wetenschappelijke Studiedagen 1994. Rotterdam: EUR.
Petersen, B. (1992). *De K-ABC en de diagnostiek van leerstoornissen. Een meta-analyse van het differentiërend vermogen van de Kaufman Assessment Battery for Children.* Nijmegen: ITS.
Procee, H. (1993). Pluralisme in (re)actie. *Comenius, 13,* (1), 56-69.
Roelandt, T., Martens, E., & Veenman, J. (1990). Achterstand van allochtonen in het onderwijs: sociaal milieu en migratie-achtergronden. *Mens en Maatschappij, 65,* (2), 103-125.
Roelandt, T., & Veenman, J. (1990). *Allochtonen van school naar werk.* 's-Gravenhage: SDU uitgeverij.
Roosens, E. (1991). Etniciteit. *Pedagogisch Tijdschrift, 16,* (2), 77-87.
Schouten, S. (1989). *Schoolloopbanen van Turken en Marokkanen in het SLVO-cohort.* Paper ORD'89. Voorburg: CBS.
Tabachnick, B., & Fidell, L. (1989). *Using Multivariate Statistics.* New York: Harper Collins Publishers.
Tacq, J. (1991). *Van probleem naar analyse. De keuze van een gepaste multivariate analysetechniek bij een sociaal-wetenschappelijke probleemstelling.* De Lier: ABC.
Tesser, P. (1989). Achterstand en cultureel verschil. Een dilemma in het onderwijsvoorrangsbeleid. *Jeugd en samenleving, 19,* (2/3), 189-201.
Tesser, P. (1993). *Rapportage minderheden 1993.* Rijswijk: SCP.
Tesser, P., Mulder, L., & van der Werf, G. (1991). *De eerste fase van de longitudinale OVB-onderzoeken. Het leerlingenonderzoek.* Nijmegen/Groningen: ITS/RION.
Tesser, P. & Vierke, H. (1990). *De schoolprestaties van allochtone leerlingen in het basisonderwijs.* Nijmegen: ITS.
Tesser, P., Werf, G. van der, Mulder, L., & Weide, M. (1989). *De beginfase van het voorrangsbeleid in het basisonderwijs. Verslag van het oriënterend onderzoek van de landelijke OVB-evaluatie.* Nijmegen/Groningen: ITS/RION.
Uiterwijk, H. (1994). *De bruikbaarheid van de Eindtoets Basisonderwijs voor allochtone leerlingen.* Arnhem: Cito.
Veenman, J. (1993). Voorzichtigheid geboden! Commentaar bij het artikel van Van 't Hof & Dronkers. *Migrantenstudies, 9,* (1), 26-30.
Wolbers, M., & Driessen, G. (1994). *Achtergronden van schoolloopbaanverschillen tussen allochtone leerlingen.* Paper ORD'94. Nijmegen: ITS.

Manuscript ontvangen 3-2-1995
Definitieve versie ontvangen 20-4-1995

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 4, pp. 363-373

Gedragstoetsen voor professionele gespreksvaardigheden: Ontwikkeling, betrouwbaarheid en validiteit

Henk T. van der Molen[1] en Wim Vermeulen[2]
Faculteit Sociale Wetenschappen, Open Universiteit en CITO, Arnhem

ABSTRACT

In this article the development, reliability and validity of behavioural tests for the assessment of professional counseling skills are discussed. The setting is the intake interview at the Community Social Services Department. The test consists of a role-play in which an actor plays the role of a client and the candidate the role of the intaker. The video-taped interview is assessed with a rating form. In this study the reliability and the validity of the behavioural test were investigated. We also compared the behavioural test rating with the criterion ratings of teachers and field experts. The internal consistency of the behavioural test appears to be satisfactory. The interrater agreement is higher than the correlation with the criterion ratings. The representativeness is good whereas the validity is corroborated.

INLEIDING

In de beroepspraktijk van de dienstverlening spelen communicatieve vaardigheden een belangrijke rol. In verschillende sectoren van het beroepsonderwijs wordt dan ook veel aandacht besteed aan het aanleren van de vaardigheden die de leerlingen later voor verschillende functies in bepaalde beroepenvelden nodig hebben (Gramsbergen & Van der Molen, 1992). Voorbeelden van deze velden zijn sociale dienstverlening, gezondheidszorg, detailhandel en bankwezen. Hoewel trainingen in gespreksvaardigheden het laatste decennium bij tal van opleidingen onderdeel uitmaken van het curriculum is het een probleem dat er nauwelijks gestandaardiseerde beoordelingsinstrumenten voorhanden zijn om deze vaardigheden adekwaat te toetsen.

Om dit probleem aan te pakken is op het Cito te Arnhem in 1988 een project gestart dat tot doel had een beoordelingsinstrument te ontwikkelen voor de toetsing van gespreksvaardigheden in één van de zojuist genoemde beroepenvelden, namelijk het veld van de sociale dienstverlening. Een opleiding die speciaal voorbereidt op dat werkveld is het Middelbaar Dienstverlenings- en Gezondheidszorg Onderwijs, afdeling Sociale Arbeid, studierichting Sociale Dienstverlening (MDGO-SA/SD). In dat onderwijs wordt als onderdeel van het vak gespreksvoering het intakegesprek geleerd omdat het in de praktijk van de dienstverlening, bijvoorbeeld bij de Gemeentelijke Sociale Dienst (GSD), een veel voorkomend gesprek is en omdat in dat gesprekstype alle belangrijke basisvaardigheden van gespreksvoering moeten worden toegepast. Concreet hebben we ons binnen het project gericht op de constructie en evaluatie van gedragstoetsen voor intakegesprekken zoals die bij de GSD dienen te worden gehouden.

Het doel van dit artikel is in de eerste plaats de ontwikkeling te beschrijven van de gedragstoetsen om te bepalen in hoeverre de onderwijsdoelstellingen met betrekking tot gespreksvoering gerealiseerd zijn. In de tweede plaats is het doel verslag te doen van uitkomsten van een empirisch onderzoek naar de betrouwbaarheid en de validiteit van de ontwikkelde toetsen.

1. Prof. dr. H.T. van der Molen is als hoogleraar psychologie verbonden aan de Open Universiteit, faculteit Sociale Wetenschappen, Postbus 2960, 6401 DL Heerlen.
2. Dr. W. Vermeulen is als wetenschappelijk medewerker werkzaam op het Cito (Instituut voor Toetsontwikkeling), Bedrijfsgroep Dienstverlening Beroepseducatie, Postbus 1034, 6801 MG Arnhem.

ONTWIKKELING VAN DE GEDRAGSTOETS

Bij de beschrijving van de gedragstoets maken we onderscheid tussen drie aspecten: 1. het simulatiespel, 2. de afnameprocedure en 3. de beoordelingsprocedure.

Het simulatiespel

In een simulatiespel wordt een deel van de sociale werkelijkheid nagebootst. In dit geval gaat het om de nabootsing van een intakegesprek tussen een cliënt en een dienstverlener van de GSD. In dit simulatiespel vervult de kandidaat de rol van dienstverlener en een acteur die van cliënt. Voor het onderzoek zijn drie rollen (casussen) ontwikkeld. De casus bestaat uit de levensbeschrijving van de cliënt en diens probleem/problemen waarmee deze voor hulp naar de GSD gaat. Om de kans te verhogen dat de casussen een vergelijkbare moeilijkheidsgraad voor de kandidaat zouden hebben, zijn ze volgens eenzelfde structuur opgezet. Om dezelfde reden is ook de instructie voor de acteurs gestandaardiseerd.

De afnameprocedure

Voorafgaand aan het simulatiespel krijgt de kandidaat een instructie. In de instructie staat dat de kandidaat de rol speelt van een bijstandsmaatschappelijk werker bij de GSD van een middelgrote gemeente. De opdracht aan de kandidaat is in maximaal 20 minuten met een cliënt een intakegesprek te voeren en de volgende doelen te realiseren: 1. het inventariseren en behandelen van materiële problemen, en 2. het inventariseren van en doorverwijzen bij immateriële problemen. De kandidaat heeft geen voorinformatie van de cliënt. Daarna wordt het simulatiespel gespeeld. Om het gedrag van de kandidaat op een systematische wijze te kunnen beoordelen, worden alle simulatiespelen op een videoband op gestandaardiseerde wijze geregistreerd.

De beoordelingsprocedure

Om een gestandaardiseerde beoordeling van de op video opgenomen gesprekken te bevorderen hebben we het beoordelingsproces in tweeën gesplitst. Voor deze twee fasen zijn twee formulieren ontwikkeld. Bij het eerste formulier, het protocolformulier, ligt het accent op de observatie van het gespreksgedrag van de kandidaat. De beoordelaar registreert tijdens het bekijken van het intakegesprek of de kandidaat de gewenste deelvaardigheden vertoont. Dit formulier bevat een aantal deelvaardigheden.

Met het tweede formulier, het beoordelingsformulier, zet de beoordelaar de registraties om in waarderingen op respectievelijk dichotoom te scoren schalen en vijfpuntsschalen. In tegenstelling tot het protocolformulier is dit formulier voor elke casus dezelfde.

Op deze manier is het proces van registratie en waardering zoveel mogelijk uit elkaar getrokken. De registratie duurt net zo lang als de dienstverlener voor het gesprek nodig heeft, het tweede deel van de beoordeling neemt ruim vijf minuten in beslag.

Concretisering van de doelstellingen: Vaardigheden en items

De doelstellingen van een intakegesprek en de vaardigheden die daarvoor nodig zijn hebben we als volgt geïnventariseerd. Voor een overzicht van de doelstellingen in een intakegesprek hebben we gebruik gemaakt van het onderscheid in doelstellingen dat Lang & Van der Molen (1991) bij hulpverlenende gespreksvoering maken. Aangezien het intakegesprek overeenkomsten vertoont met het begin van een hulpverleningsgesprek hebben we van hun inzichten gebruik gemaakt. Zij verdelen de doelstellingen van het psychologische gesprek in proces- en produktdoelen. Procesdoelen zijn doelen die de sfeer en het verloop van het gesprek faciliteren, produktdoelen zijn doelen die betrekking hebben op de effecten van de hulp (Hackney & Cormier, 1979). Lang & Van der Molen onderscheiden vier procesdoelen:

1 een relatie met een cliënt vestigen;
2 het verkennen van de problemen;
3 ordening aanbrengen in de problemen;
4 duidelijkheid aanbrengen over de verwachtingen (Lang & Van der Molen, 1991, blz. 98).

Deze vier doelen hebben we aangevuld met een vijfde, die we aan de literatuur van het probleemoplossen hebben ontleend. In het vijfde doel wordt vooral de nadruk op de beheersing van de chronologische structuur van het intakegesprek gelegd:
5 een ordelijk en systematisch verloop van het gesprek verzorgen.

Na gesprekken met deskundigen van de GSD's en docenten van het MDGO-SA/SD hebben we de produktdoelen voor het intakegesprek kunnen formuleren. In het intakegesprek gaat het in ieder geval om de volgende doelstelling:
1 inventariseren van de problemen;
Verder moet minimaal één van de volgende doelen zijn gerealiseerd:
2 informatie geven;
3 advies geven;
4 een aanvraag voor een uitkering opstellen;
5 de cliënt verwijzen.

Er zijn dus vijf proces- en vijf produktdoelstellingen in het intakegesprek te onderscheiden. Deze doelstellingen zijn als uitgangspunt genomen voor de constructie van de gedragstoets. In de literatuur hebben we gezocht naar vaardigheden die het bereiken van deze doelen bevorderen. De gevonden vaardigheden hebben we in drie clusters gegroepeerd. Elke vaardigheid is vervolgens als item in het beoordelingsschema opgenomen. Hieronder volgen de drie clusters met tussen haakjes het aantal items van het betreffende cluster:
- Gesprekstructurerende vaardigheden (20 GSV-items). Voorbeelden: 'Het kader van het gesprek aangeven', 'Het gesprek samenvatten';
- Luistervaardigheden (11 LV-items). Voorbeelden: 'Concretiserende vragen stellen', 'De cliënt betrekken bij het zoeken naar oplossingen';
- Vakinhoudelijke vaardigheden (9 VIV-items). Voorbeelden: 'Formulering correct?', 'Toelichting correct?'.

Het uitvoeren van deze vaardigheden gebeurt in een bepaalde samenhang. Daarvoor dient een gespreksmodel. Voor de keuze van een gespreksmodel hebben we aansluiting gevonden in de literatuur over hulpverlening bij Egan (1990) en Lang & Van der Molen (1991). Het gespreksmodel dat deze auteurs gebruiken, hebben wij op basis van onze ervaringen in ons onderzoek als volgt aangepast: 1. Openingsfase, 2. Probleembesprekingsfase en 3. Afsluitende fase.

In het protocolformulier hebben we de GSV-items systematisch verdeeld over de drie fasen van het gespreksmodel; de VIV-items zijn in de Probleembesprekingsfase gekoppeld aan de GSV-items. De LV-items zijn van toepassing in alle drie fasen van het gespreksmodel. De scoring van de GSV-items en de VIV-items vindt plaats op basis van vooraf gespecificeerde concrete gedragingen door middel van de antwoordcategorieën 'ja – neen' en 'plussen en minnen'. De scoring van de LV-vaardigheden geschiedt door turven van de door de beoordelaar geconstateerde voorbeelden van luistervaardig gedrag. Een deel van het protocolformulier wordt in schema 1 weergegeven.

De beoordeling van de items op het beoordelingsformulier geschiedt als volgt. Van de 20 GSV-items dienen er 17 dichotoom gescoord te worden (neen = 0, ja = 2 punten) . Bijvoorbeeld: 'Geeft de DV de tijdsduur van het gesprek aan?', 'Vat de DV het gesprek samen?'.

Bij de resterende drie items moet een beoordeling op een vijfpuntsschaal worden gegeven. Voorbeeld: 'Structureert de DV de openingsfase?'.

De beoordeling op deze vijfpuntsschaal geldt ook voor de 11 LV-items. Voorbeeld: 'Parafraseert de DV de inhoud?', 'Maakt de DV reflecterende opmerkingen?'.

De 9 VIV-items moeten objectief op een vijfpuntsschaal worden gescoord (0 - 2 - 3 - 4 - 5). Voor elke specificatie (= deelprobleem) moeten punten worden gegeven. Voorbeelden: 'Vraagt de DV adequaat door over de specificaties?', 'Geeft de DV een juiste toelichting of legt deze de procedure juist uit?'. Een deel van het beoordelingsformulier wordt weergegeven in Schema 2.

Schema 1. Deel van het protocolformulier.

CASUS 11

PROTOCOLFORMULIER		Naam kandidaat:	Naam beoordelaar:	
GESPREKSSTRUCTURERENDE VAARDIGHEDEN		VAKINHOUDELIJKE VAARDIGHEDEN	LUISTERVAARDIG-HEDEN	
OPENINGSFASE	Treedt op?			
1 Zich voorstellen	neen - ja		21 Oogcontact	.....
2 Aangeven tijd van het gesprek	neen - ja		22 Houding	.....
3 Aangeven kader van het gesprek	neen - ja		23 Kleine aanmoedigingen (verbaal/nonverbaal)	.....
4 Open beginvraag stellen	neen - ja		24 Parafraseren	.....
5 Structurering openingsfase			25 Concretiseren	.....
PROBLEEMBESPREKINGSFASE				
PROBLEEM I				
6 Identificeren specificaties:	neen - ja	32 Doorvragen specificaties?		
- hypotheek te hoog	– +	– +		
- betalingsachterst. gas-licht	– +	– +		
- idem postorderbedrijf	– +	– +		
- lesgeld Saskia	– +	– +		
7 Formuleren info/advies/ verwijzing	neen - ja	33 Form. correct?	34 Toelicht correct?	
- schuldhulpverl./Kredietbank	– +	– +	– +	
- budgettering	– +	– +	– +	
- woonkostentoeslag	– +	– +	– +	
- O&W bijdrage lesgeld Saskia	– +	– +	– +	
-	– +	– +	– +	26 Reflecteren van gevoel .....
8 Toelichten/procedures uitleggen	neen – ja			27 Accepteren van gevoel .....

METHODE

Opzet van het onderzoek naar betrouwbaarheid en validiteit

Naar aanleiding van een vooronderzoek zijn enkele wijzigingen in het instrument aangebracht en is vervolgens het hoofdonderzoek uitgevoerd. Van dit hoofdonderzoek doen we hieronder verslag.

Voor de bepaling van het aantal casussen, de vereiste grootte van de steekproeven van beoordelaars en proefpersonen hebben we gebruik gemaakt van de resultaten van het vooronderzoek door op basis van de gevonden betrouwbaarheden schattingen te doen voor gewenste betrouwbaarheden bij replicatie in het hoofdonderzoek.

Schema 2. Deel van het beoordelingsformulier.

Beoordelingsformulier

OMCIRKEL BIJ IEDER ITEM HET CIJFER VAN UW KEUZE:

Naam kandidaat:

.....

Naam beoordelaar:

.....

1 GESPREKSSTRUCTURERENDE VAARDIGHEDEN:

A OPENINGSFASE

		neen – ja
1	Stelt de DV zich voor?	0 – 2
2	Geeft de DV de tijdsduur van het gesprek aan?	0 – 2
3	Geeft de DV het kader van het gesprek aan?	0 – 2
4	Stelt de DV een open beginvraag?	0 – 2
		neen/slecht matig goed
5	Structureert de DV de openingsfase?	0 – 1 – 2 – 3 – 4

B PROBLEEMBESPREKINGSFASE

		neen – ja
	Probleem 1	
6	Identificeert de DV de specificaties?	0 – 2
7	Formuleert de DV een advies, een verwijzing of geeft deze informatie?	0 – 2
8	Geeft de DV een toelichting of legt deze een procedure uit?	0 – 2
		- - - - - - - -

Proefpersonen

In verband met het onderzoek naar de betrouwbaarheid en de validiteit zijn de gedragstoetsen afgenomen bij drie categorieën proefpersonen. De reden daarvoor was dat verwacht mocht worden dat het niveau van gespreksvoering van deze drie categorieën verschillend zou zijn. De drie categorieën waren:

– 25 eerstejaars leerlingen MDGO-SA/SD;
– 25 derdejaars leerlingen MDGO-SA/SD;
– 25 ervaren dienstverleners van GSD's.

Voor de werving van de leerlingen werden 23 verschillende scholen benaderd. Deze waren alle bereid mee te werken. Voor de onderhavige steekproeven hebben we de docenten van een achttal scholen die geografisch over het land waren verspreid gevraagd drie of zes leerlingen aselect op te geven voor het onderzoek. De dienstverleners die aan dit onderzoek hebben meegedaan zijn afkomstig van zeven GSD's en zijn allen door hun chefs als deskundige intakers gekwalificeerd.

Acteurs en casussen

Ten einde enige spreiding te krijgen in de problematiek van GSD-cliënten is besloten drie casussen voor het onderzoek te gebruiken. Drie professionele acteurs hebben elk een cliënt van een casus gespeeld. De proefpersonen hebben de drie intakegesprekken achter elkaar gevoerd. Voorafgaande aan het spelen van de casus zijn de acteurs getraind.

Toetsafname
De video-opnames zijn alle in rustige ruimten op de MDGO-scholen of bij de GSD's gemaakt. Gekozen werd voor een vaste camera-opstelling zodat de DV en face op video werd opgenomen, waarmee de uniformiteit van de opnames werd bevorderd en het verbale en non-verbale gedrag van de DV het beste kon worden beoordeeld.

Beoordelaars
De gesprekken werden door drie groepen van beoordelaars beoordeeld, namelijk:
- drie docenten MDGO-SA/SD: zij beoordeelden alle gesprekken volgens de zojuist beschreven procedure (NI);
- negen docenten MDGO-SA/SD: zij beoordeelden op de manier zoals zij dat op school deden. Dit kwam erop neer dat ze voor het gesprek een cijfer gaven dat tussen de 1 en de 10 lag (DOC);
- zes dienstverleners uit het werkveld van de GSD: zij moesten de derdejaars leerlingen voor het gesprek een waarderingscijfer geven tussen de 1 en de 10 (VELD).

De twee laatste groepen beoordelaars beoordeelden vanwege de beschikbare tijd en financiën alleen de derdejaarsleerlingen. Omdat de nieuwe beoordelingsprocedure nog onbekend was voor de betreffende groep beoordelaars werd besloten hun een beoordelingstraining te geven.

Er werd blind beoordeeld, dat wil zeggen dat de beoordelaars niet wisten of ze met eerstejaars- of derdejaarsleerlingen dan wel met dienstverleners te maken hadden. De beoordelingen vonden deels centraal, deels thuis bij de beoordelaar plaats.

Oefenmateriaal voor de scholen
Aangezien in het vooronderzoek gebleken was dat op de opleidingen MDGO het onderwijs in het intakegesprek verschillen vertoonde, is aan de deelnemende scholen geruime tijd voor de video-opnames een syllabus met een videoband verstuurd waarop een voorbeeld-intakegesprek stond. In de syllabus was beknopt een overzicht gegeven van het domein van kennis en vaardigheden dat voor het voeren van een intakegesprek relevant is. Op die manier kon de docent ervoor zorgen dat de leerlingen voor de afname zoveel mogelijk een gelijke uitgangspositie hadden.

Betrouwbaarheid
Voor het onderzoek naar de betrouwbaarheid van de gedragstoetsen zijn de interne consistentie en de beoordelaarsovereenstemming berekend. Voor deze laatste coëfficiënt hebben we gebruik gemaakt van de coëfficiënt van Gower (G_{xy}). De coëfficiënt van Gower (Gower, 1971) is een maat voor absolute overeenstemming tussen twee beoordelaars. Deze coëfficiënt kan waarden aannemen in het interval tussen 0 en 1. Hofstee en Zegers (1991) stellen voor de G_{xy} te transformeren naar "$2G_{xy} - 1$" , zodat het interval van de getransformeerde G_{xy} tussen -1 en +1 komt te liggen. Deze suggestie hebben we overgenomen.

Validiteit
Bij het onderzoek naar de validiteit van de gedragstoets is getracht op de volgende drie vragen een antwoord te vinden. Ten eerste hebben we met behulp van een enquête aan de betrokken docenten en werkvelddeskundigen gevraagd wat hun mening was over de representativiteit van de items uit de gedragstoets.

Ten tweede werd nagegaan of het oordeel van via de nieuwe procedure getrainde beoordelaars samenhing met het oordeel van experts. Ten slotte is voor de drie betrokken groepen, waarvan verwacht mag worden dat die in prestatieniveau verschillen, onderzocht of de gemiddelde groepsprestaties inderdaad van elkaar verschillen.

RESULTATEN

Betrouwbaarheid

Interne consistentie

In Tabel 1 staat de interne consistentie van de gedragstoets per casus, apart en gesommeerd over de drie beoordelaars-NI en gebaseerd op het totaal van de drie steekproeven van proefpersonen.

Zowel voor elke beoordelaar als voor de drie casussen worden hoge tot zeer hoge coëfficiënten gevonden. Ook worden zowel bij de drie clusters van vaardigheden als bij de drie afzonderlijke steekproeven hoge tot zeer hoge interne consistenties gevonden. Op grond hiervan hebben we geconcludeerd dat de interne consistentie van de gedragstoetsen bevredigend is.

Interbeoordelaarsovereenstemming

In Tabel 2 volgt voor de gedragstoets de beoordelaarsovereenstemming, berekend over de drie beoordelaars-NI, per casus en steekproef.

De beoordelaarsovereenstemming voor de procedure bij de gedragstoets is hoog te noemen en ligt in dezelfde orde van grootte als de gevonden beoordelaarsovereenstemming in het vooronderzoek (Vermeulen, 1993). Bij een vergelijking tussen de drie steekproeven blijkt de beoordelaarsovereenstemming voor de dienstverleners van de GSD's wat lager te liggen, vermoedelijk door "restriction of range". De kleine spreiding bij de dienstverleners wordt waarschijnlijk

Tabel 1. Interne consistentie (Cronbach's alpha) van de gedragstoetsen voor alle proefpersonen (n = 75), per casus, apart en gesommeerd over drie beoordelaars-NI.

	CASUS I	CASUS II	CASUS III
1	0.91	0.89	0.90
2	0.92	0.90	0.91
3	0.93	0.91	0.92
TOTAAL	0.94	0.92	0.93

1, 2, 3 = beoordelaar 1, 2 en 3
TOTAAL = Sommatie van de scores over de drie beoordelaars

Tabel 2. Beoordelaarsovereenstemming voor de beoordelingsprocedure bij de gedragstoets, per casus, berekend over de drie beoordelaars, over eerstejaars, derdejaars leerlingen MDGO-SA/SD en dienstverleners van GSD's (elk n = 25) afzonderlijk en totaal.

	GOWER*		
CASUS	I	II	III
EERSTEJAARS MDGO-SA/SD	0.90	0.90	0.90
DERDEJAARS MDGO-SA/SD	0.88	0.86	0.92
DIENSTVERLENERS GSD	0.84	0.86	0.86
TOTAAL	0.88	0.88	0.90

GOWER* = De getransformeerde $G_{xy} = 2G_{xy} - 1$

Tabel 3. Gemiddelde beoordelaarsovereenstemming voor het docentoordeel en het werkveldoordeel over de drie casussen, gebaseerd op derdejaars leerlingen.MDGO-SA/SD (n = 25).

	GOWER*	
	DOC (k = 9)	VELD (k = 6)
GEMIDDELD	0.77	0.79

DOC = Docentoordeel
VELD = Werkveldoordeel
k = Aantal beoordelaars
GOWER* = De getransformeerde $G_{xy} = 2G_{xy} - 1$

veroorzaakt door het feit dat de dienstverleners allen goed op de gedragstoets presteren waardoor de scores dicht bij elkaar komen te liggen. Als geheel genomen kan de beoordelaarsovereenstemming van de gedragstoetsen goed genoemd worden.

Om inzicht te hebben in de betrouwbaarheid van het docentoordeel en het werkveldoordeel geven we in onderstaande tabel de gemiddelde beoordelaarsovereenstemming voor beide oordelen over de drie casussen, gebaseerd op derdejaars leerlingen MDGO-SA/SD (n = 25):

De tabel laat zien dat met de coëfficiënt van Gower voor beide oordelen een vrijwel gelijke, aanzienlijke beoordelaarsovereenstemming wordt bereikt. Voor een vergelijking van de beoordelaarsovereenstemming tussen enerzijds de beoordeling volgens de procedure bij de gedragstoets en anderzijds het docent- en het werkveldoordeel verwijzen we naar Tabel 2 en 3. Daaruit blijkt dat met de beoordeling volgens de procedure bij de gedragstoets de hoogste beoordelaarsovereenstemming wordt behaald.

Ten aanzien van de betrouwbaarheid van de gedragstoets kan de conclusie luiden dat we ten eerste zowel voor de interne consistentie als voor de beoordelaarsovereenstemming hoge waarden hebben gevonden. Ten tweede kan worden vastgesteld dat de beoordelaarsovereenstemming voor de gedragstoetsen beduidend hoger is dan die voor het docent- en het werkveldoordeel.

VALIDITEIT

Representativiteit

Aan de docenten en werkvelddeskundigen (n = 18) bij de GSD die bij het hoofdonderzoek waren betrokken, is een enquête verzonden. Daarin is hen de vraag voorgelegd in hoeverre naar hun mening de beoordeelde kennis en vaardigheden uit de gedragstoets, zoals geoperationaliseerd in de 40 items, een representatieve afspiegeling vormen van de onderhavige beroepsvaardigheid, het voeren van een intakegesprek. Uit de resultaten van 16 geretourneerde enquêteformulieren kon worden geconcludeerd dat de items uit de gedragstoets als voldoende representatief werden beschouwd.

Samenhang tussen de scores op de gedragstoetsen en oordelen van docenten en werkvelddeskundigen

Vervolgens hebben we in het onderzoek naar de validiteit het oordeel van de docenten in het MDGO en dat van de werkvelddeskundigen van de GSD over de intakegesprekken als criteriumbeoordeling genomen. De reden daarvoor is dat zowel de docent gespreksvoering in het MDGO als de dienstverlener van de GSD als expert op het gebied van het voeren van intakegesprekken kunnen worden beschouwd. Verwacht mag worden dat experts het beste in staat zijn de beoogde beroepsvaardigheid te beoordelen. Een ondersteuning voor de validiteit van de

Tabel 4. Gemiddelde Pearson-productmomentcorrelaties tussen de gemiddelde scores van de gedragstoetsbeoordelingen (k = 3) en de gemiddelde cijfers van het docentoordeel (k = 9) en die van het werkveldoordeel (k = 6), over drie casussen, bij derdejaars leerlingen MDGO-SA/SD (n = 25).

	NI – VELD	NI – DOC	VELD – DOC
GEMIDD	0.79	0.86	0.87

NI = Beoordeling volgens de procedure bij de gedragstoets
DOC = Docentoordeel
VELD = Werkveldoordeel
k = Aantal beoordelaars

gedragstoets wordt verkregen wanneer aangetoond wordt dat bij de beoordeling van de intakegesprekken de scores van de kandidaten, die verkregen zijn met de beoordelingsprocedure van de gedragstoets, substantieel samenhangen met de cijfers verkregen via de criteriumbeoordelingen van de docenten en de werkvelddeskundigen. De correlaties staan weergegeven in tabel 4.

Uit Tabel 4 blijkt dat de beoordeling volgens de procedure bij de gedragstoets in hoge mate samenhangt met zowel het docentoordeel als met het werkveldoordeel. De drie oordelen meten in sterke mate dezelfde beroepsvaardigheid. Deze uitkomst beschouwen we als een ondersteuning van de validiteit.

Verschillen tussen groepen

Het derde deel van het onderzoek naar de validiteit heeft betrekking op de vraag of met de gedragstoets verschillen kunnen worden aangetoond tussen de gemiddelde prestatie van groepen die geacht worden in prestatieniveau te verschillen (Hofstee, 1971; Thorndike, 1982). In Tabel 5 geven we voor elke casus de gemiddelden en standaarddeviaties van de gedragstoetsbeoordelingen voor de eerstejaars- en derdejaarsleerlingen MDGO-SA/SD alsmede de dienstverleners van de GSD (steeds n = 25).

Teneinde te achterhalen of de steekproeven als factor Niveau en de Casussen als factor van invloed zijn op de scores is een variantie-analyse uitgevoerd. Het resultaat van deze analyse staat in Tabel 6 vermeld.

Uit Tabel 5 en 6 blijkt dat het Niveau als factor significant van invloed is op de score op het intakegesprek. Tussen de eerste- en derdejaars leerlingen MDGO-SA/SD is geen significant verschil gevonden (Tukey test 5.46, n.s.). Het verschil tussen de derdejaarsleerlingen MDGO-

Tabel 5. Gemiddelden en standaarddeviaties van de gedragstoetsbeoordelingen, gebaseerd op de drie groepen proefpersonen (steeds n = 25), per casus.

	EERSTEJAARS		DERDEJAARS		DV	
	GEMIDD	SD	GEMIDD	SD	GEMIDD	SD
CASUS I	65.17	17.38	70.69	17.86	80.61	11.80
CASUS II	62.76	13.52	67.13	13.41	83.07	9.61
CASUS III	53.56	12.88	59.40	14.74	74.73	14.43

EERSTEJAARS = Eerstejaars leerlingen MDGO-SA/SD
DERDEJAARS = Derdejaars leerlingen MDGO-SA/SD
DV = Dienstverleners van de GSD

Tabel 6. Variantie-analyse op de factoren Niveau (de steekproeven eerstejaars, derdejaars leerlingen MDGO-SA/SD en werkvelddeskundigen, steeds n = 25) en Casus (de drie casussen).

	DF	F	Pr > F
NIVEAU	2	35.86	0.0001
CASUS	2	10.23	0.0001
ERROR	4	0.44	0.78

SA/SD en de dienstverleners van de GSD blijkt wel significant te zijn (Tukey test 5.46, $p < 0.05$).

Verder laat de tabel zien dat ook de Casus als factor van invloed is op de score. Het maakt derhalve wel wat uit welke casus de kandidaat moet behandelen. Deze uitkomst staat in de literatuur bekend als casusspecificiteit (Van der Vleuten & Swanson, 1990).

Deze twee gegevens laten de volgende conclusies toe. Aangezien met de beoordelingsprocedure bij de gedragstoets het verschil in niveau van de drie steekproeven significant kan worden aangetoond, betekent dit ten eerste een bijdrage aan de validiteit van de gedragstoets. De tweede conclusie, dat de casussen die hetzelfde construct meten significant van elkaar in moeilijkheidsgraad kunnen verschillen, wordt in de literatuur bevestigd. Opgaven in dit soort settingen waar sprake is van simulatie zijn specifiek: de prestatie van een kandidaat op een opgave doet geen uitspraak over de te verwachten prestatie op andere opgaven die hetzelfde vermogen bedoelen te meten (Kraan & Crijnen, 1987).

Conclusies en discussie

De algemene conclusie van het onderzoek luidt dat het mogelijk is intakegesprekken betrouwbaar en valide te meten. Dit wordt bevestigd door de volgende bevindingen. In het onderzoek hebben we de ontwikkelde gedragstoets getoetst op de aspecten betrouwbaarheid en validiteit. Ten aanzien van de betrouwbaarheid is ten eerste gebleken dat gedragstoets in hoge mate intern consistent is. Ten tweede is de betrouwbaarheid nagegaan door de beoordelaarsovereenstemming van de beoordeling bij de gedragstoets te berekenen. Deze bleek hoog te zijn. Een vergelijking ten aanzien van de beoordelaarsovereenstemming tussen enerzijds de beoordelingsprocedure bij de gedragstoets en anderzijds het docentoordeel en het oordeel van het werkveld valt in het voordeel van de gedragstoets uit.

Wat de validiteit van de gedragstoets betreft blijkt ten eerste de representativiteit bevredigend te zijn. Ten tweede hebben we de samenhang berekend tussen de beoordelingsprocedure bij de gedragstoets en de twee criteriumoordelen: het docentoordeel en het oordeel van het werkveld. Deze samenhang bleek substantieel . Ten slotte bleek het mogelijk de verwachte verschillen tussen de scores van in niveau verschillende groepen aan te tonen. Op grond van deze uitkomsten concluderen we dat deze deelonderzoeken steun verlenen aan de validiteit van de gedragstoets.

Een kanttekening is bij bovenstaande conclusie op zijn plaats. De beoordelaarstraining kan namelijk aanleiding geven tot discussie. Deze is namelijk wel aan de beoordelaars met de gedragstoets en niet aan de andere twee groepen beoordelaars gegeven. Deze training zou de gunstige uitslag voor de gedragstoets beïnvloed kunnen hebben. Als reactie daarop kan worden gezegd dat het in het onderzoek in eerste instantie om de waardebepaling van de gedragstoets ging en niet om een vergelijking tussen de gedragstoets en de huidige beoordelingssituatie in geoptimaliseerde vorm. Om die reden hebben we een training voor de docenten en de werkveldeskundigen achterwege gelaten.

Ten slotte willen we nog iets zeggen over het probleem van de generalisatie van de huidige

resultaten naar andere situaties waarin gespreksvoeringsonderwijs plaatsvindt. Op verschillende terreinen wordt met succes onderzoek gedaan naar het beoordelen van professionele gespreksvaardigheden: in de opleiding van huisartsen (Van der Vleuten & Van Luyk, 1988), op het gebied van de hulpverlening in de geestelijke gezondheidszorg (Bögels, 1994) en bij de rechtenstudie (Smilde, 1992). Gedragstoetsen spelen daarin een belangrijke rol. Het nadeel van gedragstoetsen is evenwel dat ze moeilijk efficiënt zijn uit te voeren: de afname en beoordeling nemen veel tijd in beslag, terwijl standaardisatie lastig te bereiken is. In dit verband is het van belang dat onderzocht wordt of er alternatieve methoden voorhanden zijn om professionele gespreksvoering even betrouwbaar en valide te meten als met gedragstoetsen het geval is (Smit, 1994).

LITERATUUR

Bögels, S.M. (1994). *Teaching and assessing diagnostic interviewing skills. An application to the mental health field.* Maastricht: Universitaire Pers Maastricht.

Egan, G. (1990). *The skilled helper. A systematic approach to effective helping.* (4th Edition). California: Brooks/Cole Publishing Company.

Gower, J.C. (1971). A general coefficient of similarity and some of its properties. *Biometrics, 27,* 857-874.

Gramsbergen-Hoogland, Y. & Molen, H.T. van der (1992). *Gesprekken in organisaties.* Groningen: Wolters-Noordhoff.

Hackney, H. & Cormier, L.S. (1979). *Counseling Strategies and Objectives.* Englewood Cliffs, N.J.: Prentice-Hall.

Hofstee, W.K.B. (1971). Begripsvalidatie van studietoetsen. *Nederlands Tijdschrift voor de Psychologie, 26,* 491-500.

Hofstee, W.K.B. & Zegers, F.E. (1991). Idiographic correlations: modeling judgments of agreement between school grades. *Tijdschrift voor Onderwijsresearch, 16,* 331-336.

Kraan, H.F. & Crijnen, A.A.M. (1987). *The Maastricht History-Taking and Advice Checklist: Studies of Instrumental Utility.* Proefschrift. Amsterdam: Lundbeck.

Lang, G. & Molen, H.T. van der (1991). *Psychologische gespreksvoering. Een basis voor hulpverlening.* Baarn: H. Nelissen.

Smilde, M. Het gesprek tussen advocaat en cliënt. Analyse en aanbevelingen. *De Psycholoog, juni,* 255-259.

Smit, G. (1994). De beoordeling van professionele gespreksvaardigheden. *De Psycholoog, juli/augustus,* 266-269.

Thorndike, R.L. (1982) *Applied Psychometrics.* Boston: Houghton Mifflin Company.

Vermeulen, W. (1993). *Toetsing van communicatieve vaardigheden. Constructie en evaluatie van gedragstoetsen voor professionele gespreksvaardigheden.* Proefschrift. Arnhem: Cito.

Vleuten, C.P.M. van der & Swanson, D.B. (1990). Assessment of clinical skills with standardized patients: state of the art. *Teaching and Learning in Medicine,* 2, 58-76.

Manuscript ontvangen 16-11-1994
Definitieve versie aanvaard 10-4-1995

Tijdschrift voor Onderwijsresearch, 20 (1995), nr. 4, pp. 374-376

Boekbespreking

M.K. van der Heijden: *Consistentie van aanpakgedrag. Een procesdiagnostisch onderzoek naar acht aspecten van hoofdrekenen.* Proefschrift R.U. Leiden. Lisse: Swets & Zeitlinger, 1993.

In "Leren en instructie" van Boekaerts en Simons, dat evenals deze dissertatie van Van der Heijden in 1993 is verschenen, besteden de auteurs in hoofdstuk 4 veel aandacht aan meta-cognitieve vaardigheden, vergelijkbaar met de begrippen aanpak (gedragingen) en leerstrategieën. Hoe belangrijk Boekaerts en Simons deze vaardigheden voor een optimale voortschrijding van leerprocessen achten, moge uit de volgende uitspraken blijken:

- Het aanleren van cognitieve en meta-cognitieve strategieën is een van de belangrijkste nieuwe taken van de leerkracht (hoofdstuk 10).
- De effectiviteit van het onderwijs wordt opzienbarend vergroot wanneer meta-cognitieve kennis en vaardigheden worden aangeleerd (hoofdstuk 10).
- Kinderen met leermoeilijkheden beschikken over minder meta-cognitieve vaardigheden dan leerlingen die de betreffende moeilijkheden niet hebben (p. 94).
- Meta-cognitieve vaardigheden zijn belangrijke voorspellers van leerprestaties (p. 94).

Alleen al deze uitspraken wijzen op de noodzaak om bij diverse schoolvakken na te gaan hoe leerlingen schooltaken aanpakken, welke strategieën tot optimale leerresultaten leiden, of aanpakgedragingen binnen een taakdomein consistent zijn, etc. Welnu, Maarten van der Heijden heeft zulks ten aanzien van het rekenen gedaan. Preciezer gezegd, en overeenkomstig de ondertitel van het proefschrift, heeft hij een procesdiagnostisch onderzoek naar acht aspecten van hoofdrekenen gedaan.

Onder aanpak(gedrag) van een leerling bij een bepaalde (leer)taak of opgave verstaat Van der Heijden de wijze waarop die leerling met de taak of opgave omgaat. Het aanpakgedrag wordt beschreven in termen van aspecten ervan, namelijk:

1. Aandacht; bij dit aanpakaspect wordt nagegaan of de leerling tekorten vertoont bij het oplezen van de opgaven (plus in plaats van min of omgekeerd) en (of) bij het opschrijven van het antwoord (de leerling schrijft bijvoorbeeld 61 in plaats van 16).
2. Oriëntering; bij oriënterend handelen wordt de opgave verkend met het doel de meest efficiënte oplossing te vinden. Daarom spreekt Van der Heijden bij dit aspect van handig rekenen. Daarvan is bijvoorbeeld sprake wanneer een leerling bij de opgave 15+28-15 =? de termen 15 tegen elkaar laat wegvallen.
3. Oplossingsprocedures; in dit kader worden onder andere de rijg- en splitsmethode onderscheiden. Bij de rijgmethode wordt het eerste getal heel gelaten en vermeerderd of verminderd met de tientallen en eenheden afzonderlijk van het tweede getal, bijvoorbeeld 35+17= / 35+10=45 / 45+7=52 . Bij de splitsmethode worden beide getallen gesplitst in tientallen en eenheden en worden vervolgens de tientallen respectievelijk de eenheden bij elkaar opgeteld of van elkaar afgetrokken.
4. Automatisering; bij dit aanpakaspect gaat het om het niveau van de telhandelingen (op de vingers, hardop of in het hoofd) en om verkorting van de uitgevoerde deelhandelingen.
5. Flexibiliteit; hiervan spreekt Van der Heijden wanneer een leerling de oplossingsprocedure aanpast aan de kenmerken van de opgave. Zo is bij het aftrekken zonder tiental-passering de splitsmethode handig, maar bij aftrekken met tiental-passering de rijgmethode.
6. Inzichtelijkheid; van inzichtfouten is bijvoorbeeld sprake wanneer eenheden bij tiental-passering verwisseld worden.
7. Bewustheid; de bewustheid van het handelingsverloop blijkt uit de mate waarin een leerling naar aanleiding van een retrospectievraag het door hem gevolgde handelingsverloop adequaat kan verwoorden.
8. Controle; onder dit aanpakaspect worden bijvoorbeeld fouten-analyses en correctiehandelingen gerekend.

Van der Heijden heeft het onderzoek naar deze acht aanpakaspecten van hoofdrekenen uitgevoerd in het kader van het handelingspsychologisch interactiemodel; een driedelig model dat bestaat uit de componenten situatie, persoon en aanpak. Dit model geeft aan dat wanneer een leerling geconfronteerd wordt met een taaksituatie er uit de leerling/situatie-interactie een bepaalde aanpak resulteert. Alleen al uit de benaming van enkele aanpakaspecten, zoals oriëntering en controle, blijkt dat de auteur gebruik maakt van het begrippenapparaat van Gal'perin.

Om de verschillende aanpakaspecten van het hoofdrekenen vast te stellen, is gebruik gemaakt van procesdiagnostiek, ook wel kwalitatieve diagnostiek genoemd. Om de genoemde aspecten te diagnostiseren heeft Van der Heijden de Leidse Diagnostische Rekenaanpak Test (LDRT) ontwikkeld. De test, die bij 174 eind-vierde-groep-basisschoolleerlingen van gemiddeld acht jaar en drie maanden individueel is afgenomen, bestaat uit hoofdrekenopgaven, namelijk optellen en aftrekken tot 100.

De belangrijkste resultaten van het onderzoek zijn:

- De onderzochte leerlingen vertonen in het algemeen een hoge mate van consistentie in aanpakgedrag. Vooral bij de mate van automatisering en het proceduregebruik ligt de consistentie hoog.
- De variabiliteit in aanpakgedrag wordt in het algemeen meer door de individuele verschillen tussen de leerlingen bepaald dan door verschillen (in type en moeilijkheidsgraad) tussen de opgaven.
- Bewustheid, controle en de mate van automatisering van deelhandelingen blijken de centrale aspecten van het rekenaanpakgedrag te vormen.
- De rekenprestaties kunnen goed voorspeld worden op grond van de onderscheiden aspecten van het aanpakgedrag gezamenlijk.
- De rekenaanpak hangt niet samen met de verbale intelligentie (woordenschat).
- De zwakkere hoofdrekenprestaties van meisjes ten opzichte van die van jongens kunnen verklaard worden door de verschillen in aanpak tussen de geslachten.

Omdat de aspecten bewustheid en controle, die als centrale aspecten van de rekenaanpak naar voren zijn gekomen, in het algemeen niet als niet-taakspecifiek worden beschouwd, acht Van der Heijden een eventuele consistentie in aanpakgedrag bij *verschillende* leertaken mogelijk. Zulks geldt bijvoorbeeld ten aanzien van de taakdomeinen rekenen en bewegingsonderwijs. Ook bij het leren van motorische vaardigheden, zoals in het bewegingsonderwijs, spelen de aanpakaspecten bewustheid (in de vorm van gedragcentrering, zelfsturing en structureren) en controle (in de vorm van fouten-analyses) een belangrijke rol bij de voortschrijding van het leerproces. Mede op grond van het feit dat ook de andere door Van der Heijden onderscheiden aanpakaspecten gelijk zijn of overeenkomen met die welke bij het leren van sportieve vaardigheden zijn onderscheiden, suggereert Pijning (1994, p. 350-352) dat er met betrekking tot deze leertaken sprake is van consistentie in aanpakgedrag.

Van der Heijden heeft aangetoond dat bij het uit het hoofd optellen en aftrekken tot 100 de meeste eind-vierde-groep-basisschoolleerlingen een 'slaafse, onkritische en inflexibele aanpak volgen' (derde stelling). Hij verwijt het onderwijs dan ook dat het 'bij leerlingen niet tot de gewoonte heeft geleid om eerst de opgave te analyseren, alvorens met rekenhandelingen te beginnen' (p. 6). Om de hoofdrekenprestaties van leerlingen te verhogen zal het onderwijs zich moeten richten op 'optimalisering van aspecten van het rekenaanpakgedrag, zoals oriënteren, controleren en analyseren van fouten' (p. 278).

Gezien het feit dat er steeds meer voor wordt gepleit de ontwikkeling van meta-cognitieve vaardigheden tot onderwijsdoel te maken (Boekaerts en Simons, 1992, p. 228), is dit proefschrift, dat ook in handelseditie is verschenen een schot in de roos. Zoals uit het voorgaande moge blijken gaat het hier om een gedegen onderzoek naar de aspecten van de aanpak bij het hoofdrekenen. Andere interessante thema's die in dit boek op heldere wijze aan de orde komen zijn:

- de operationalisatie van het begrip aanpakgedrag, waardoor het begrip ook in andere taakdomeinen kan worden toegepast;

- het feit dat het begrip aanpakgedrag een brugfunctie vervult tussen persoonlijkheid en leren.

Gezien de belangrijke didactische consequenties van deze studie is het te hopen dat een en ander snel tot de pabo's doordringt!

Literatuur

Boekaerts, M. & Simons, P.R.J. (1993). *Leren en instructie. Psychologie van de leerling en het leerproces.* Assen: Dekker v.d. Vegt.
Pijning, H.F. (1993). *Motoriek en leerpsychologie.* Haarlem: De Vriesborch.
Pijning, H.F. (1994). Consistentie van aanpakgedrag. *Lichamelijke Opvoeding,* 350-352.

H.F. Pijning

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres
Dr. M. Mulder,
Faculteit Toegepaste Onderwijskunde,
Universiteit Twente,
Postbus 217, 7500 AE Enschede,
tel. 053-893652

Jrg. 19, nr. 1 1995

Inhoudsopgave

Lezing opening ORD 95 378

Onderwijsresearch bedreigd?

(Tekst lezing bij opening ORD95 te Groningen)

Verscheidenheid

Namens de VOR wil ik de organisatie van de ORD 95 graag complimenteren met de organisatie van het programma dat de komende twee dagen gepresenteerd wordt. Het bijna 300 pagina's dikke congresboek toont een groot scala aan thema's. Mij valt vooral de rijke verscheidenheid op die al zo vele jaren kenmerkend is voor het onderwijsonderzoek in ons land. Dat heeft naast iets aantrekkelijks ook een schaduwzijde: het leidt ook gauw tot versnippering en gebrek aan coherentie. Een ordening in thema's die voor een deel samenvallen met de VOR divisies kan de nodige samenhang zeker bevorderen. Kennisname van het aanbod per thema laat zien dat toch ook daar nog van een grote verscheidenheid sprake is. Op dit punt maakt het onderwijsonderzoek nog niet veel vorderingen.

In 1975 kenmerkte De Groot de onderwijsresearch in de empirische onderwijskunde onder meer als: "een wetenschap waarin de research-produktie wordt gekenmerkt door divergentie, ook qua uitgangspunten en basisbegrippen, en door een overmaat aan 'kleingoed' zonder veel verband; een wetenschap die een relatief lage status heeft, bij andere wetenschappen en bij het publiek." Twintig jaar later zijn er helaas weinig tekenen die er op duiden dat deze kenmerken niet langer van toepassing zijn. Voor een belangwekkend en informatief overzicht kan ik wijzen op de inventarisatie van zwaartepunten in het Nederlandse onderwijsonderzoek, die Van Gennip en Pelkmans (1994) opgesteld hebben voor de Overlegcommissie Verkenningen. Het zou een goede zaak zijn als de komst van de onderzoekschool ICO (Interuniversitair Centrum voor Onderwijskundig Onderzoek) ook enige verbetering zou kunnen aanbrengen in de samenhang van wetenschappelijke kennis over het onderwijs.

De organiserende instelling heeft bij elke ORD de mogelijkheid een eigen thematiek te kiezen. Het spreekt welhaast vanzelf dat Groningen, of moet ik zeggen GION, koos voor 'Evaluatie van onderwijseffectiviteit'. Het thema is van actueel belang en kent inmiddels een belangwekkende samenhangende kennisgroei. Zeker in een tijd waarin de autonomie van scholen vergroot wordt, is het nodig aandacht te besteden aan de kwaliteit van het onderwijs dat in een grotere vrijheid gegeven wordt. Het ontwikkelen van een goed instrumentarium om scholen en bestuurders zicht te geven op hetgeen in scholen gepresteerd wordt is onderwijskundig en onderwijspolitiek van groot belang.

Ongerustheid

Bert Creemers opent het congresboek met een aardige karakteristiek van onderwijsonderzoek. Ik citeer zijn eerste zin: "Er is altijd wel iets in het onderwijsonderzoek aan de hand: er wordt geëvalueerd, er is een verkenningscommissie, er moeten Centres of Excellence of onderzoekscholen worden opgericht, of er is ergens ruzie over."

Ik zou er nog aan toe willen voegen: of er is ergens ongerustheid over. Dat geldt zeker nu we weten dat de Tweede kamer akkoord is gegaan met de voornemens van het kabinet ten aanzien van de toekomstige regeling van de onderwijsverzorging. Het voornemen op het gebied van onderwijsonderzoek is het fundamenteel alsmede het strategisch, evaluatief en toegepast onderzoek onder te brengen bij de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). Een overweging hierbij is dat het onderbrengen van deze typen onderzoek bij één organisatie de onderlinge relatie zou moeten versterken. Voorwaarde is dat de NWO een apart programma voor onderwijsonderzoek opbouwt. Het opstellen van een dergelijk programma gebeurt met behulp van een programmaraad van deskundigen.

Voor het kortlopend onderzoek worden financiële middelen ondergebracht bij de landelijke pedagogische centra (VSLPC), die voor de aanbesteding bij onderzoeksinstituten zorg zullen dragen. Het Ministerie krijgt eveneens de beschiking over een budget voor kortlopend onder-

zoek. Hiervoor is in totaal dertig procent van het huidige SVO-onderzoeksbudget beschikbaar, gelijk te verdelen over beide groeperingen.

Het Instituut voor Onderzoek van het Onderwijs (SVO) gaat dus verdwijnen. Ik vrees dat daarmede de positie van onderwijsonderzoek in onzeker vaarwater terecht gekomen is. Het is dringend noodzakelijk dat er zo gauw mogelijk doorzichtige procedures voor de programmering van onderwijsonderzoek bij NWO komen. Ook dient duidelijk te worden op welke wijze de financiering - NWO kent tot nu toe een van de SVO afwijkende financiering - geregeld wordt. Over de effecten van de nieuwe verdeling van onderzoeksgelden tasten we nog in het duister. Ik maak mij grote zorgen of we erin kunnen slagen de opgebouwde expertise in de onderzoeksinstituten te behouden. Vanuit de VOR zullen we er alles aan proberen te doen zo spoedig mogelijk duidelijkheid te verkrijgen in de genoemde kwesties. We hebben in maart 1995 in een brief aan de Staatssecretaris onze ongerustheid gemeld. Zij heeft daarop uitvoerig geantwoord, overigens zonder daarbij haar voornemens te wijzigen. Nu de besluitvorming in de Tweede Kamer is afgerond lijkt het voor de VOR verstandig nader overleg te voeren met de directie van de NWO.

De laatste tijd heb ik me herhaaldelijk afgevraagd waaraan het onderwijsonderzoek het heeft verdiend dat zoveel veranderd wordt. Het merkwaardige is dat er ook internationaal sprake is van een verzwakking van het onderwijsonderzoek. Ook internationaal is op het gebied van onderwijsonderzoek reden tot ongerustheid.

Op de 'Annual Meeting' van de American Educational Research Association (AERA) dit jaar in San Francisco woonde ik onder meer een symposium bij dat onder leiding stond van Torsten Husèn over de positie van Research & Development (R&D). In een presentatie van Harnqvist werd opgemerkt dat de publieke ondersteuning voor R&D in veel landen afneemt. Hij noemde als oorzaken:

1. De problemen staan te ver af van de praktijk.
2. De resultaten hebben vaak het karakter van open deuren en lijken triviaal.
3. De informatie komt niet terecht bij de juiste doelgroepen.
4. Zelfs indien mogelijk, blijft toepasbaarheid van gegevens beperkt.

Deze situatie leidt tot een dilemma: Zorg dragen voor een hoge kwaliteit bij de wetenschappelijke kennisproduktie versus verhoging van de effectiviteit bij het verbeteren van de onderwijspraktijk.

Harnqvist wil de oplossing zoeken in een benadering waarbij sprake is van meer samenwerking met de praktijk. Onderzoek heeft dan wat minder distantie. Wardekker (1986) heeft dat aangeduid met het onderscheid tussen het waarnemersperspectief en het deelnemersperspectief. Harnqvist presenteert een nieuwe variant van actie-onderzoek. Hij spreekt van 'learning communities'. Het betreft een interactiemodel tussen onderzoekers en betrokkenen. Geprobeerd moet worden samen het probleem te formuleren, er voldoende tijd voor te nemen, uit te gaan van interdisciplinariteit, indien nodig verschillende paradigma's toe te staan en de verschillende niveaus van betrokkenen in het stelsel erbij te betrekken. Dit alles moet in relatie staan met het vigerende beleid en met de praktijk voor zover de groep daarop invloed kan uitoefenen.

Het lijkt mij van groot belang dat onderwijsonderzoekers in internationaal verband zich beraden over deze problematiek. Het verbeteren van het imago van onderwijsonderzoek vraagt om een alert pr-beleid en goede communicatie met de afnemers.

Europese Vereniging

In dit verband is het plezierig te vermelden dat de oprichting van de Europese Vereniging van Onderwijsonderzoek nu een feit is. In de conferentie in Bath in Engeland in september 1995 zal de eerste algemene ledenvergadering van de EERA (European Educational Research Association) plaats vinden. Het gaat vooral om een vereniging van verenigingen. Tot nu toe zijn 9 verenigingen uit even zoveel landen aangesloten en 10 onderzoeksinstituten. Er is een liaison met een nieuw Engelstalig tijdschrift en de VOR leden kunnen zich voor slechts 50 gulden daarop abonneren. Vooralsnog heeft men afgezien van een indeling in de gebruikelijke divisies.

Men hoopt dat netwerken zullen ontstaan tussen onderzoekers uit verschillende landen, die daarvoor een eigen thematiek kiezen.

Ik heb inmiddels vastgesteld dat ook in Bath weer veel Nederlandse onderzoekers present zullen zijn. Laten we proberen zo veel mogelijk te profiteren van de internationale uitwisseling.

De kwaliteit van het Nederlandse onderwijsonderzoek kan een internationale vergelijking goed doorstaan. Helaas heeft men in het buitenland niet altijd zo veel interesse in hetgeen elders wordt gedaan. Dat geldt vooral in de Verenigde Staten. Op de Amerikaanse ORD, de AERA, waar zo'n 10.000 bezoekers komen, was voor de Nederlandse presentaties bedroevend weinig belangstelling. Ik ben benieuwd hoe dat in Europa zal zijn.

Ik ben ook erg nieuwsgierig naar de resultaten van een nieuw fenomeen in de kwaliteitszorg rondom onderwijsonderzoek: de onderzoekvisitatie. Dit jaar heeft een internationaal samengestelde commissie onder leiding van Knoers al het eerste geldstroom onderzoek en het daarmee verbonden overige onderzoek beoordeeld. Op dit moment is de rapportage nog niet beschikbaar. Na de positieve ervaringen met de onderwijsvisitatie is het van groot belang dat ook het onderzoek aan een stelsel van kwaliteitszorg wordt onderworpen.

Met grote interesse kijk ik uit naar de bevindingen van de verkenningscommissie onderwijsonderzoek. Naar verluidt rapporteert deze in september.
Ik ben vooral benieuwd wat die rapportage nog kan betekenen voor de verdere voortgang van het onderwijsonderzoek onder de vleugels van de NWO.

Missie

De VOR heeft al enige jaren zijn missie verbreed tot al het werk dat onderwijskundigen in hun beroepsuitoefening doen. De VOR verenigt dus meer dan alleen onderwijsonderzoekers. Dat roept wel een aantal vragen op omtrent het karakter van het verenigingsblad en over de activiteiten die de VOR als totaliteit moet of kan organiseren. Voorlopig ligt het accent op de activiteiten die de divisies ontplooien. Het is denkbaar dat de VOR in de toekomst meer op de maatschappij gerichte activiteiten gaat ontwikkelen. In de jaren vijftig en zestig was de WVO (Werkgemeenschap voor Onderwijsvernieuwing) erg actief in het organiseren van nationale conferenties over vraagstukken inzake onderwijsvernieuwing. Inmiddels zorgen overheid, verzorgingsinstellingen en nascholingsinstituten voor een levendig conferentie aanbod. Zou hier niet een taak voor de VOR liggen? Er is immers nogal wat te verbeteren aan de communicatie tussen onderwijskunde en de samenleving. Door het naar buiten brengen van belangwekkende informatie over onderwijspolitieke thema's kan aan verbetering van de interactie worden gewerkt.

Buiten het onderwijsonderzoek blijkt het goed te gaan met de onderwijskunde en de onderwijskundigen. De opleidingen lijken hun studentenaantallen op peil te kunnen houden. Er is voorts weinig sprake van werkeloosheid onder afgestudeerden.

In toenemende mate verschijnen nieuwe overzichtswerken, waardoor oude paradepaardjes als 'Beknopte didaxologie' (laatste editie 1982) en 'Onderwijs: bestel en beleid' (laatste versie 1986) eindelijk gelijkwaardige opvolgers krijgen. Wat nog ontbreekt is een nieuwe 'Inleiding in de onderwijskunde' (de laatste editie onder redactie van Lagerweij & Vos is van 1987).

Het vak heeft bestaansrecht verworven en is niet meer weg te denken uit het onderwijsbestel. In het landelijk overleg in de Kamer onderwijskunde van de VSNU is met instemming gereageerd op een voorstel uit Utrecht om in het voorjaar van 1996 een conferentie te organiseren over het onderwerp: de grondslagen van de onderwijskunde. Centrale thematiek is de vraag naar de wetenschappelijkheid van de onderwijskunde. De initiatiefnemers Lennart Vriens en Rob de Jong menen dat de onderwijskunde als wetenschap in een overgangsfase verkeert. De oorsprong ligt volgens hen in een pre-theoretische fase, waarbij object en taak gedefinieerd werden vanuit extern-gebonden vraagstukken. Helaas heeft de onderwijskunde te weinig gereflecteerd op haar object, waardoor, aldus de initiatiefnemers, een enorme fragmentatie is ontstaan. Zij vragen zich af of er nog wel een duidelijk identificeerbare onderwijskunde als wetenschap bestaat. Er is bijvoorbeeld nauwelijks geschikte literatuur te vinden voor het onderwijs over de grondslagen van deze wetenschap. De literatuur wordt grotendeels nog ontleend aan de basisdisciplines.

Kortom alle reden voor een nationale bezinning op onderwijskunde als wetenschap. Het lijkt mij ook in het licht van de door mij hiervoor genoemde positie van onderwijsonderzoek een goed initiatief en ik hoop dat het initiatief ondersteuning krijgt vanuit alle centra waar onderwijsvraagstukken bestudeerd worden.

Ik kom nu toe aan het verlenen van een erelidmaatschap van onze vereniging. Vorig jaar werd dit voor het eerst in de geschiedenis van de VOR verleend aan A.D. de Groot en E. Warries. Vandaag doet het mij veel genoegen het erelidmaatschap te mogen uitreiken aan Fons Knoers. De VOR is bijzonder verheugd dat hij het erelidmaatschap van de vereniging heeft willen aanvaarden. Het betreft een persoon die bijzondere verdiensten heeft gehad voor het onderwijs en voor de ontwikkeling van de onderwijskunde in Nederland.

Prof.dr N.A.J.Lagerweij
Vakgroep onderwijskunde
Universiteit Utrecht
Voorzitter VOR

Literatuur

Gennip, J. Van, Pelkmans, A. (1994). *Zwaartepunten in het Nederlandse onderwijsonderzoek.* Inventarisatie voor de Overlegcommissie Verkenningen. Nijmegen: ITS.

Groot, A.D. de (1975). *Ontwikkelingslijnen in de Nederlandse onderwijsresearch.* Amsterdam: RITP

Wardekker, W.L. (1986). *Wetenschapstradities en onderwijsvernieuwing.* Amsterdam: VU Uitgeverij.

Artikelen

Deregulering en de marktpositie van scholen
H.W.C. van Amelsvoort en B. Witziers 115

Ontwikkelingen in het omgaan met literaire complexiteit
E. Andringa 271

Huiswerk maken bij de televisie of bij geluidsmedia
J.W.J. Beentjes, C.M. Koolstra en T.H.A. van der Voort 178

Het ontwerpen van tests met verschillende optimaliteitscriteria
M.P.F. Berger en W.J. van der Linden 79

A self-evaluation procedure for schools using multilevel modelling
R. J. Bosker and J. Scheerens 154

Aandachtsstrategieën bij het reviseren in een vreemde taal: de invloed van de taakomgeving
H. Broekkamp en H. van den Bergh 245

Leerstijlenonderzoek gerepliceerd; de samenhang tussen leerstijlen en intelligentie
V.V. Busato, F.J. Prins, C. Hamaker en K.H. Visser 332

Het relatieve belang van sociaal milieu en etnische herkomst voor de verklaring van onderwijsachterstanden
G.W.J.M. Driessen 341

Leesstrategieën van anderstalige en Nederlandse leerlingen in het voortgezet onderwijs
H. Hacquebord 256

Denominatie en effectiviteit van schoolbesturen
R. Hofman, W. Hoeben en H. Guldemond 63

De evaluatie van het basisonderwijs door de CEB
F.G.J. Janssen 3

Tijdsbesteding in combinatieklassen en enkelvoudige klassen. Een descriptief onderzoek naar besteding van de beschikbare tijd op basisscholen
M. Kral, M. Voeten en W. Nederlof 300

De obrengsten van het basisonderwijs: een repliek
W.J. van der Linden en M.A. Zwarts 34

Robustness of judgments in evaluation research
W.J. van der Linden and M.A. Zwarts 13

Classroom composition and individual achievement. Effects of classroom composition and teacher goals in Dutch elementary education
H. Luyten and A. van der Hoeven-van Doornum 42

Gedragstoetsen voor professionele gespreksvaardigheden: Ontwikkeling, betrouwbaarheid en validiteit
H. T. van der Molen en W. Vermeulen 363

Proces en startegie in een psycholinguïstisch model van schrijven en lezen
E. van der Pool en C. van Wijk 200

Leren en onderwijzen van strategieën in het talenonderwijs
G. Rijlaarsdam 197

Strategische competentie in het talenonderwijs
G. Rijlaarsdam, R. Oostdam en P. Bimmel 215

Inleiding: Nuances in de autonomie
J. Scheerens 97

Studiesucces van buitenlandse studenten: Studievoortgang van buitenlands studenten in het technisch-wetenschappelijk onderwijs geanalyseerd
H.J. Schmid, A. Verweij, W. Jochems en J. Snippe 165

De rol van strategieën en metacognitie bij het leren van woorden in het moedertaal- en vreemde-talenonderwijs
C. Schouten-van Parreren, K. de Glopper en M. van Dalen-Kapteijns 229

Beoordeling van het motorische niveau en de schrijfmotoriek van basisschoolleerlingen door leerkrachten
B. C.M. Smits-Engelsman, G. P. van Galen en C. G.J. Michels 285

De taakinhoud, de taakomvang en de taakbelasting van schoolleiders in het basisonderwijs
W.G.R. Stoel 133

De opbrengsten van het basisonderwijs volgens de CEB: enkele kanttekeningen bij de gevolgde normeringsprocedures
J.M. Wijnstra 28

Is strategische hulp effectief in een aardrijkskunde-computerles?
T. van der Zijpp 318

Boekbesprekingen

D. Dolmans: *How students learn in a problem-based curriculum* (J. Vermunt) 191

Geert ten Dam, Edith van Eck en Monique Volman: *Onderwijs en sekse. Een verkenning van researchprogramma's* (C.A. Klaassen) 194

M.K. van der Heijden: *Consistentie van aanpakgedrag. Een procesdiagnostisch onderzoek naar acht aspecten van hoofdrekenen* (H.F. Pijning) 374

Notities en Commentaren

Neuzen tellen?
H. Blok 96

Een 'Boekbespreking als boemerang: Blok of het wetenschappelijk forum'
T. Mooij 93

Redactioneel

Van de redactie
R. Voeten 1

Redactionele medewerkers

In 1995 werd redactionele medewerking verleend door:

P. Been	Rijksuniversiteit Groningen
J.J. Beishuizen	Rijksuniversiteit Leiden
H.A. Bosma	Rijksuniversiteit Groningen
T.F. Brandsma	Universiteit Twente Enschede
J. Cohen-Schotanus	Rijksuniversiteit Groningen
M. van Dijk	Advies Raad Onderwijs Utrecht
D. van Dongen	IVA Tilburg
J.W.J.M. Driessen	ITS Nijmegen
J. Dronkers	SCO-Kohnstamm Instituut Amsterdam
E. Elshout-Mohr	SCO-Kohnstamm Instituut Amsterdam
W. Fase	Erasmus Universiteit Roterdam
G.P. van Galen	Katholieke Universiteit Nijmegen
K. de Glopper	SCO-Kohnstamm Instituut Amsterdam
E.G. Harskamp	GION Groningen
W.Th.J.G. Hoeben	GION Groningen
W.K.B. Hofstee	Rijksuniversiteit Groningen
J.F.M.J. van Hout	Universiteit van Amsterdam
G.G.H. Jansen	Rijksuniversiteit Groningen
P. Jungbluth	ITS Nijmegen
L.J.Th. van der Kamp	Rijksuniversiteit Leiden
H.P.M. Krammer	Universiteit Twente Enschede
G. Lang	Rijksunversiteit Groningen
J.M.G. Leune	Erasmusuniversiteit Rotterdam
A.F.M. Nieuwenhuis	STOAS Wageningen
A.M. Overmaat	SCO-Kohnstamm Instituut Amsterdam
J.L. Peschar	Rijksuniversiteit Groningen
S.J. Pijl	GION Groningen
G. Reezigt	GION Groningen
J.F. Schröder	Rijksuniversiteit Groningen
P.R.J. Simons	Katholieke Unversiteit Nijmegen
H.H. Tillema	Rijksuniversiteit Leiden
G.J. Tillekens	Rijksuniversiteit Groningen
S.A.M. Veenman	Katholieke universiteit Nijmegen
J.D.H.M. Vermunt	Katholieke Universiteit Brabant
L. Verschaffel	Katholieke Universiteit Leuven
T.H.A. van der Voort	Rijsuniversiteit Leiden
R.F.A. Wierstra	Rijksuniversiteit Utrecht
W.H.F.W. Wijnen	Rijksuniversiteit Limburg