

Retentie-effecten van een nascholingsprogramma voor effectieve instructie en klasmanagement*

S. Veenman en J. Raemaekers**

Samenvatting

In deze studie worden de lange-termijn effecten van een nascholingsprogramma voor effectieve instructie en klasmanagement beschreven. In twee eerdere studies werd ingegaan op de korte-termijn effecten. De eerste evaluatiestudie werd uitgevoerd in het schooljaar 1986/87, de tweede in 1989/90. In de tweede evaluatiestudie werd ook het effect van coaching beproefd. Eind 1992 werden de lange-termijn of retentie-effecten van het nascholingsprogramma onderzocht. Dit geschiedde door middel van een 'untreated control group design' met een voormeting, een nameting en een retentiemeting. De gegevens betreffende de vergelijking tussen de experimentele groep en de controlegroep lieten zien dat het nascholingsprogramma een significant effect had op de taakgerichtheid van de leerlingen en op aspecten van het instructie- en klasmanagementgedrag van leerkrachten. Een duurzaam effect van coaching werd niet gevonden. Ook werden geen significante verschillen gevonden tussen de experimentele groep die vijf jaar geleden het programma had gevolgd en de experimentele groep die twee jaar geleden het programma had gevolgd. Tussen de nameting en de retentiemeting werden geen significante verschillen aangetroffen. De vooruitgang die direct geboekt werd na de training werd voor de meeste instructie- en klasmanagementvaardigheden vastgehouden, een verdere groei of toename trad niet op. Een effect van het nascholingspro-

gramma op de leerprestaties van de leerlingen werd niet gevonden.

1 Achtergrond

In het basisonderwijs neemt de combinatieklas een prominente plaats in. Volgens de Commissie Evaluatie Basisonderwijs (1994) geeft 53% van de basisschoolleerkrachten les in een combinatieklas. In een combinatieklas geeft de leerkracht les aan twee of meer groepen tegelijkertijd. Op grond van observaties en interviews met leerkrachten uit combinatieklassen werden vijf probleemgebieden geïdentificeerd bij het geven van onderwijs in combinatieklassen, namelijk 1) het efficiënt gebruik van de leertijd, 2) het geven van effectieve instructie, 3) doelmatig klasmanagement, 4) het organiseren van zelfstandig leren, en 5) het schoolbeleid met betrekking tot combinatieklassen (Veenman, Lem, Voeten, Winkelmolen & Lassche, 1986). Op basis van deze vijf probleemgebieden werd een teamgerichte nascholingscursus ontwikkeld, getiteld *Omgaan met combinatieklassen* (Veenman, Lem & Nijssen, 1988)¹. De inhoud van de nascholing was gebaseerd op recente empirische inzichten over goed en succesvol onderwijzen. In twee evaluatiestudies werden de korte-termijn effecten van het nascholingsprogramma vastgesteld. De opzet en de resultaten van deze studies zijn in twee eerdere artikelen in dit tijdschrift beschreven (Veenman, Lem & Roelofs, 1990; Roelofs, Veenman & Raemaekers, 1993) en worden beknopt in paragraaf 4 samengevat. Dit artikel richt zich op de retentie- of lange-termijn effecten van het nascholingsprogramma.

* Dit onderzoek werd uitgevoerd met behulp van een subsidie van het Instituut voor Onderzoek van het Onderwijs (SVO-project 1111).

** Dank zijn wij verschuldigd aan de volgende personen voor hun bijdragen: M. Voeten, C. van den Berg, M. van Moorsel, C. Motshagen, D. Jacobs, N. Hofs en de deelnemende schoolleiders en leerkrachten.

2 Training en transfer

In studies waarin getracht wordt de uitkomsten van trainingen vast te stellen, wordt opgemerkt dat het onderzoek zich tot nu toe weinig gelegen heeft laten liggen aan de transfer-effecten van trainingen (Baldwin & Ford, 1988; Broad & Newstrom, 1992; Simons, 1992). Veel nascholingscursussen leiden niet tot een toepassing van de geleerde inzichten en vaardigheden in de klas als werksituatie (Van Tulder, 1992). De overdracht van het geleerde in een cursus of training naar de werksituatie wordt in de literatuur aangeduid met het begrip 'transfer'. Wanneer deze overdracht is geslaagd is sprake van positieve transfer. Transfer treedt op als de lerende in staat is het geleerde van de trainingssituatie te generaliseren naar de werksituatie en het aldaar gedurende langere tijd toepast. Transfer houdt ook in dat met toenemende oefening op de werkplek het niveau van de toegepaste kennis en vaardigheden na verloop van tijd hoger ligt dan aan het einde van de training.

Gezien de geringe transfer-effecten voor allerlei soorten van trainingen, nemen Broad en Newstrom (1992) aan dat ongeveer 50% van alle trainingsinhouden één jaar na afloop van de training nog toegepast worden. Om de condities voor het optreden van transfer-effecten te versterken stellen zij de vorming van 'transfer partnerships' voor. Dit zijn samenwerkingsverbanden tussen de drie participanten die belang hebben bij de training: de lerenden (de trainees), de trainers (de ontwerpers en de instructeurs van de training) en de managers (de leidinggevendenden in de organisatie die verantwoordelijk zijn voor de toepassing van de trainingsinhouden op de werkplek). Iedere partner kan vóór, tijdens en na de training een aantal maatregelen nemen om het proces van transfer te bevorderen. Door de tijdsdimensie (vóór, tijdens, na de training) te combineren met de rol-dimensie van de betrokken partners (manager, trainer, lerende) hebben Broad en Newstrom (1992) een matrix ontworpen met transferbevorderende maatregelen. Een aantal van de door hen aangereikte suggesties is verwerkt in de opzet en de uitvoering van de nascholing *Omgaan met Combinatieklassen*. Zo werden tijdens een intake-gesprek de leerkrachten op de hoogte gebracht van de doelen, de inhouden, de werkwijzen van de nascholing. Ook werd

gewezen op het nut of de bruikbaarheids-waarde van de inhouden van de nascholing. Zowel de docenten die de nascholing verzorgden als de leerkrachten die de nascholing volgden werden betrokken bij de opzet en uitwerking van het programma. De leerkrachten die aan de nascholing deelnamen hadden allemaal de intentie om hun onderwijs in combinatieklassen te verbeteren. Volgens Den Ouden (1992) vormen de intenties van de lerenden een cruciaal concept bij het aanbrengen en het verklaren van gedragsverandering na een training of opleiding. Gegevens over het feitelijke gedrag van leerkrachten - verzameld via observaties in de klas - werden tijdens de nascholing gebruikt om verbeteringsvoorstellen te formuleren in het licht van de aangeboden informatie. De nascholing was systematisch opgezet en uitgevoerd volgens de door Joyce en Showers (1995) beschreven componenten van effectieve training. In teamverband werden gezamenlijke voorstellen voor verbetering geformuleerd. Oefensituaties om de nieuwe vaardigheden toe te passen werden gecreëerd; tijdens de training via gesimuleerde condities en tijdens het lesgeven via praktijkgeoriënteerde opdrachten. Leerkrachten kregen de mogelijkheid om hun ervaringen met de toepassing van de nieuwe vaardigheden te bespreken met collega's. In gezamenlijk overleg werden lesplannen voorbereid en geëvalueerd. Om transfer te bevorderen werd bovendien voorzien in coaching.

3 Coaching

Studies naar de relatie tussen vormgeving en effecten van nascholing laten zien dat nascholing leidt tot transfer als naast de training ook sprake is van begeleiding of coaching in de klas (Bennett, 1987; Joyce & Showers, 1995). Bij programma's zonder coaching zijn de transfer-effecten minimaal. Coaching wordt door Joyce en Showers (1980) beschreven als: "Hands-on, in-classroom assistance with the transfer and application of skills to the classroom". Aan coaching kunnen vier functies worden onderscheiden: (1) het geven van steun en vertrouwen aan de leerkracht die poogt nieuwe kennis en vaardigheden in zijn klas toe te passen, (2) het geven van inhoudelijke terugkoppeling

over de toepassing van nieuwe kennis en vaardigheden in de klas, (3) het analyseren van de nieuwe kennis en vaardigheden op hun gebruikswaarde in de klas, en (4) het afstemmen van de nieuwe kennis en vaardigheden op de leerlingen.

4 Korte-termijn effecten van het nascholingsprogramma

Zoals gezegd werden in twee evaluatiestudies de korte-termijn effecten van de nascholing vastgesteld. In de eerste evaluatiestudie waren medewerkers van de Vakgroep Onderwijskunde betrokken bij de opzet en uitvoering van de training. In het schooljaar 1986/87 namen 41 leerkrachten van 8 schoolteams aan de nascholing deel (combinaties 2/3 tot en met 7/8). Uit deze groep werden vervolgens 17 leerkrachten geselecteerd voor observaties in de klas (experimentele groep). Negen leerkrachten uit 6 vergelijkbare scholen vormden de controlegroep (zie Tabel 1, Appendix). Zij namen niet aan de nascholing deel. De leerkrachten uit de beide groepen werden direct vóór en na de nascholing tijdens twee rekenlessen en twee lees/taallessen geobserveerd.

In de tweede evaluatiestudie werd op vijf plaatsen in Nederland de nascholing verzorgd door PABO-docenten en schoolbegeleiders. Deze ontvingen een voorscholing om te garanderen dat het programma op eenzelfde wijze werd uitgevoerd als in de eerste studie. In het schooljaar 1989/90 namen 89 leerkrachten van 12 scholen deel aan de nascholing. Uit deze groep werden 28 leerkrachten (groep 3 tot en met 8) geselecteerd voor lesobservaties (experimentele groep). Van deze groep ontvingen 18 leerkrachten naast de nascholing ook nog coaching door de schoolbegeleiders. De controlegroep werd gevormd door 14 leerkrachten van 6 vergelijkbare scholen (zie Tabel 1).

De nascholing omvatte vijf tot zeven bijeenkomsten van elk drie uur, die om de twee weken plaatsvonden. De weken tussen de bijeenkomsten waren bestemd voor het uitvoeren van praktijkopdrachten. De gegevens van de observaties werden naar de leerkrachten teruggekoppeld zodat zij op grond van deze gegevens en de gegeven theorie verbeteringsplannen konden formuleren. Deze gegevens hadden

vooral betrekking op het taakgerichte gedrag van de leerlingen en op de instructie- en klasmanagementvaardigheden van de leerkrachten.

Uit de resultaten van de eerste evaluatiestudie bleek dat de nascholing effectief was. Er werd een significant effect vastgesteld ten aanzien van het taakgerichte gedrag van de leerlingen en ten aanzien van een aantal instructie- en klasmanagementvaardigheden (Veenman, Lem & Roelofs, 1990). Ook in de tweede evaluatiestudie werd een significant effect van de nascholing vastgesteld op het taakgerichte gedrag van de leerlingen en op het instructie- en klasmanagementgedrag van de leerkrachten. Twee effecten voor coaching werden gevonden, namelijk voor de organisatie van de instructie en voor het omgaan met storingen in de klas. In klassen van gecoachte leerkrachten werd een sterkere vooruitgang in taakgerichte leertijd geconstateerd dan in klassen van niet-gecoachte leerkrachten (Roelofs, Veenman & Raemaekers, 1993).

5 Onderzoeksvragen

Deze studie richt zich op de lange-termijn of retentie-effecten van het nascholingsprogramma *Omgaan met Combinatieklassen*. Het onderzoek kent de volgende vraagstellingen: 1) Passen de leerkrachten die de nascholing in de schooljaren 1986/87 en 1989/90 hebben gevolgd de gewenste instructie- en klasmanagementvaardigheden nog steeds toe? 2) Heeft de nascholing een blijvend effect op het taakgerichte gedrag van de leerlingen? 3) Zijn de effecten van de nascholing groter voor die leerkrachten die ook coaching hebben ontvangen? 4) Heeft de nascholing effect op de leerprestaties van de leerlingen?

6 Opzet

De toetsing van de effecten van nascholing en coaching op het instructie- en klasmanagementgedrag van de leerkrachten geschiedde door middel van een 'untreated control group design' met een voormeting, een nameting en een retentiemeting. Het onderzoek kende een controlegeconditie ($n = 11$) en twee experimen-

tele condities: niet-gecoachte leerkrachten ($n = 10$) en gecochte leerkrachten ($n = 8$). Ter vaststelling van het effect van de nascholing en coaching op de leerprestaties van de leerlingen werd gebruik gemaakt van een 'posttest-only design with nonequivalent groups', waarbij de experimentele groep bestond uit 26 klassen en de controlegroep uit 14 klassen.

6.1 Onderzoeksgroep

Oorspronkelijk was het de bedoeling om alleen de scholen uit de tweede evaluatiestudie te benaderen voor deelname aan de retentiestudie. In deze studie werden vóór en na de nascholing niet alleen observaties uitgevoerd, maar werden ook schooltoetsen afgenomen om het effect van de nascholing op de leerprestaties te kunnen vaststellen. In de eerste evaluatiestudie waren geen schooltoetsen afgenomen. De animo van scholen en leerkrachten om aan de retentiestudie deel te nemen was gering. Door de late start van het onderzoek hadden enkele scholen hun prioriteiten anders gelegd. Sommige leerkrachten waren vertrokken. Een aantal leerkrachten wilde vanwege de taakbelasting slechts deelnemen aan een beperkt aspect van de retentiestudie (bijvoorbeeld wel aan de observatiestudie, maar niet aan de afname van schooltoetsen). Een aantal scholen merkte op dat zij vroeger aan de studie hadden deelgenomen omdat zij daar baat bij hadden in de vorm van nascholing; de retentiestudie gaf hun geen profijt. Door deze experimentele uitval, inherent aan elke longitudinale studie, werd besloten ook de leerkrachten te benaderen die de nascholing in het schooljaar 1986/87 hadden gevolgd. Deze uitbreiding gaf de mogelijkheid ook het tijdseffect te onderzoeken: maakten leerkrachten die vanaf 1986/87 met de inhoud van de nascholing werkten, in vergelijking met de leerkrachten die vanaf 1989/90 hiermee werkten, meer gebruik van de gewenste instructie- en klasmanagementvaardigheden?

De samenstelling van de onderzoeksgroepen voor de eerste en tweede evaluatiestudie is beschreven in Veenman, Lem en Roelofs (1990) en in Roelofs, Veenman en Raemaekers (1993). Het navolgende handelt over de onderzoeksgroep van de retentiestudie.

Van de 17 getrainde leerkrachten die deelnamen aan het observatie-onderzoek van de eerste evaluatiestudie uit 1986/87 bleken 5 jaar

later 8 leerkrachten bereid opnieuw deel te nemen aan de observaties voor de retentiestudie. Van de 10 getrainde maar niet-gecoachte leerkrachten die deelnamen aan het observatie-onderzoek van de tweede evaluatiestudie uit 1989/90 bleken 2 jaar later 2 leerkrachten bereid deel te nemen aan de retentiestudie. Aldus namen 10 getrainde maar niet-gecoachte leerkrachten deel aan de retentiestudie. Van de 18 gecochte leerkrachten uit de tweede evaluatiestudie namen 8 leerkrachten deel aan de retentiestudie. In totaal namen 18 getrainde leerkrachten (gecoacht of niet-gecoacht) aan de retentiestudie deel (zie Tabel 1). Deze leerkrachten gaven les in de groepen 3 tot en met 8 en waren afkomstig van 9 scholen.

De controlegroep bestond uit 11 leerkrachten die verbonden waren aan 7 scholen met vergelijkbare sociaal-economische en geografische achtergronden als de scholen uit de experimentele groepen (6 van de 9 leerkrachten uit de controlegroep van de eerste evaluatiestudie en 5 van de 14 leerkrachten uit de controlegroep van de tweede evaluatiestudie).

Ter controle op mogelijke zelf-selectie-effecten werden allereerst de scores op de voormeting betreffende het taakgerichte gedrag van de leerlingen en de instructie- en klasmanagementvaardigheden van de aan de retentiestudie deelnemende klassen en leerkrachten vergeleken met die van de niet aan de retentiestudie deelnemende klassen en leerkrachten. Uit de gegevens blijkt dat de scores van de experimentele groep die deelnam aan de retentiestudie ($n = 18$) op 5%-niveau niet significant afweken van de scores van de experimentele groep die niet deelnam aan de retentiestudie ($n = 27$). Voor de controlegroep werd slechts één significant verschil gevonden. De leerkrachten uit de controlegroep die deelnamen aan de retentiestudie ($n = 11$) scoorden significant lager dan de leerkrachten uit de controlegroep die niet deelnamen aan de retentiestudie ($n = 12$) op een van de vijf subschalen betreffende het management- en instructiegedrag, namelijk 'afstemming van de instructie'. Vervolgens werden de scores op de nameting van de aan de retentie-meting deelnemende en niet-deelnemende leerkrachten met elkaar vergeleken. Niet uitgesloten kon worden dat de aan de retentie-studie deelnemende leerkrachten zich onderscheiden door enthousiasme en inzet tij-

dens de training, wat zou kunnen resulteren in hogere scores op de nameting. Dit bleek niet het geval. Voor de nameting werden geen significante verschillen tussen de experimentele en de controlegroep aangetroffen. Op grond van deze vergelijkingen op de voor- en name-ting werd geconcludeerd dat er nauwelijks sprake was van zelf-selectie op de voor het onderzoek relevante variabelen.

Het aantal leerkrachten dat in de retentiestudie bereid was om schooltoetsen af te nemen, was groter dan het aantal leerkrachten dat bereid was om observatoren in de klas toe te laten. Van de 12 klassen in de retentiestudie met getrainde maar niet-gecoachte leerkrachten, hadden 7 leerkrachten aan de eerste evaluatiestudie deelgenomen en 5 aan de tweede evaluatiestudie. Alle klassen in de retentiestudie met gecoachte leerkrachten ($n = 14$) namen ook deel aan de tweede evaluatiestudie (in de eerste evaluatiestudie vormde coaching geen experimentele conditie). De controlegroep bestond uit 14 leerkrachten (7 uit de eerste en 7 uit de tweede evaluatiestudie). Zie voor een overzicht Tabel 1.

Hoewel alle leerkrachten die deelnamen aan het observatie-onderzoek uit de eerste en de tweede evaluatiestudie benaderd werden, was het aantal leerkrachten dat bereid was tot deelname aan de retentiestudie niet indrukwekkend (40% van de totale experimentele groep en 48% van de totale controlegroep uit beide evaluatiestudies). Zoals reeds eerder opgemerkt kwam dit doordat de leerkrachten en scholen geen onmiddellijk profijt hadden van de retentiestudie. Andere argumenten om van deelname af te zien waren: andere prioriteiten, overbelasting, ziekte en leerkrachten die de school hadden verlaten.

6.2 Instrumentarium

De meetinstrumenten die gebruikt werden om het effect van de nascholing of training vast te stellen op het taakgerichte gedrag van de leerlingen en het instructie- en klasmanagementgedrag van de leerkrachten waren identiek aan die van de eerste en tweede evaluatiestudie. Gebruikt werden een observatie-categorieën-systeem en een beoordelingschaal. Om het effect op de leerlingprestaties vast te stellen werd gebruik gemaakt van gestandaardiseerde toetsen.

6.2.1 Het observatie-instrument

Voor de registratie van het taakgerichte gedrag van de leerlingen en de instructie- en klasmanagementvaardigheden van de leerkracht werd gebruik gemaakt van het observatie-instrument COMMIT. Hiermee werd een schatting verkregen van de tijd die door de leerlingen en leerkrachten aan verschillende activiteiten en settings wordt besteed. De observatoren registreerden: 1) de taakgerichte leertijd van leerlingen, 2) de groepeeringsvorm (setting) waarin de leerlingen werkten, 3) de activiteit van de leerkracht en 4) de doelgroep van de leerkracht. Gebruikt werd een specifieke vorm van 'time-sampling', namelijk de 'predominant activity sampling'-procedure (Tyler, 1979). De procedure verliep als volgt: eerst werden de activiteiten van leerling één en de leerkracht geobserveerd (duur: 7 seconden), waarna de geobserveerde gedragingen werden gecodeerd in de vier boven beschreven rubrieken van het observatie-instrument (duur: 13 seconden). Vervolgens werden de activiteiten van leerling twee en de leerkracht geobserveerd en gecodeerd, waarna de procedure werd herhaald tot elke leerling uit de klas was geobserveerd en gecodeerd. Hierbij werden de leerlingen in een vooraf vastgestelde volgorde geobserveerd.

Voorafgaande aan de observaties volgden drie observatoren een observatie-training van ongeveer 40 uur. Bij de training werd gebruik gemaakt van video-fragmenten en lesobservaties in reële klassituaties. De training werd verzorgd door een ervaren trainer uit de tweede evaluatiestudie die op zijn beurt getraind was door een ervaren trainer uit de eerste evaluatiestudie. Dit ter waarborging van een identieke observatie-procedure in de drie studies. De tussen-observator-betrouwbaarheid voor de verschillende observatie-categorieën, geschat met behulp van variantie-analyse (Winer, 1971), varieerde van .70 tot 1.00 (mediaan .97), met uitzondering van de categorie 'doelgroep hele klas' (.25). Deze categorie werd niet in de verdere analyse betrokken.

6.2.2 De beoordelingschaal

Met behulp van de Management- en Instructie Schaal (MIS) werden meer specifieke gegevens verzameld over de kwaliteit van de management- en instructievaardigheden van de leerkracht. De MIS was gebaseerd op onder-

zoek naar effectief managementgedrag van leerkrachten (Evertson, Emmer, Sanford & Clements, 1983; Veenman, Lem, Voeten, Winkelmolten & Lassche, 1986) en op onderzoek naar kenmerken van effectieve instructie (Rosenshine & Stevens, 1986; Good, Grouws & Ebmeier, 1983). De 31 items van de schaal representeerden vijf subschalen: 1) instructievaardigheden, 2) organisatie van de instructie, 3) gebruik van ruimte en materialen, 4) de afstemming van de instructie op de leerlingen en 5) het omgaan met storingen. Deze items werden ook gebruikt in de eerste en tweede evaluatiestudie. De alpha-coëfficiënten voor deze subschalen varieerden van .66 tot .91 (zie Tabel 3). De tussen-observator-betrouwbaarheid varieerde van .63 tot .81 (mediaan .74).

6.2.3 Schooltoetsen

Om het effect van de nascholing op de leerprestaties vast te stellen werden in de groepen 3 tot en met 8 drie toetsen afgenomen: een voor technisch lezen (de Een-Minut-Test (EMT) van Brus en Voeten), een voor begrijpend lezen (CITO-Begrijpend lezen, Begrippentest Aarnoutse, Begrijpend lezen Aarnoutse en Hoofdgedachtetest Aarnoutse) en een voor rekenen (Tempo-Test-Rekenen van De Vos). Voor alle groepen was de EMT de enige identieke test. Van de Tempo-Test-Rekenen werd de gehele toets afgenomen in de groepen 5 tot en met 8, in de groepen 3 en 4 een gedeelte van de test. Voor begrijpend lezen bestonden geen toetsen die voor alle groepen van de basisschool geschikt waren. Hier werden verschillende toetsen gebruikt; elke test trachtte uitgaande van een bepaald leerjaar de vaardigheid van begrijpend lezen te meten.

6.3 Dataverzameling en -verwerking

De voormeting van de eerste evaluatiestudie vond plaats in de maanden november-december 1986, de nameting in de maanden mei - juni 1987. Gedurende de tweede evaluatiestudie vond de voormeting plaats in de maanden november-december 1989 en de nameting in de maanden mei-juni 1990. De meting voor de retentiestudie vond plaats in de periode mei-juli 1992, vijf jaar na de nameting van de eerste evaluatiestudie en twee jaar na de nameting van de tweede evaluatiestudie.

Alle leerkrachten die deelnamen aan de re-

tentiestudie werden geobserveerd tijdens een reken- en een lees/taalles. Alle observatiegegevens die werden verkregen met behulp van de COMMIT hadden per meting betrekking op de gemiddelde tijdsbesteding per klas of per leerkracht, berekend over de twee geobserveerde lessen. Per observatiecategorie resulteerde een tijdschatting die werd uitgedrukt in minuten (geobserveerd werden lessen van 40 minuten). De observatiegegevens van de MIS werden per subschaal omgezet in somscores (berekend door de itemscores voor elk van de vijf subschalen te combineren). Deze scores hadden eveneens betrekking op de gemiddelden van twee lessen.

Bij de toetsing van verschillen tussen de experimentele groep en de controlegroep en tussen de experimentele groepen onderling werd een significantieniveau van 5% aangehouden (eenzijdig). Voor de observatiegegevens vormde de klas of leerkracht de analyse-eenheid; voor de gegevens van de schooltoetsen werd de leerling als analyse-eenheid genomen. Voor de toetsing van verschillen werd gebruik gemaakt van het programma MANOVA van SPSS-X. Voor meer gedetailleerde gegevens over de opzet, het instrumentarium en de dataverzameling wordt verwezen naar Raemaekers en Veenman (1994).

7 Resultaten

Ter toetsing van het effect van het nascholingsprogramma op de taakgerichte leertijd van de leerlingen en het instructie- en klasmanagementgedrag van de leerkracht werden covariantie-analyses (ANCOVA) uitgevoerd met de scores op de voormeting als covariabelen. De scores van de nameting en de retentiemeting vormden de afhankelijke variabelen, met 'treatment' (training of geen training) als onafhankelijke variabele. Uitgevoerd werden drie vergelijkingen: (1) experimentele groep versus controlegroep, (2) coaching versus geen coaching en (3) de nascholingsgroep van 1986/87 versus de nascholingsgroep van 1989/90. Opgemerkt moet worden dat alleen de scores van de niet-gecoachte leerkrachten uit de tweede evaluatiestudie vergeleken konden worden met die van de leerkrachten uit de eerste evaluatie-

studie, omdat coaching geen experimentele conditie vormde in de eerste evaluatiestudie.

Allereerst werd getoetst of er op de voormeting verschillen bestonden tussen de experimentele groep (getrainde leerkrachten met of zonder coaching) en de controlegroep. Er werden geen significante aanvangsverschillen gevonden tussen beide groepen. Tussen de experimentele groep met en zonder coaching werd op de voormeting alleen voor de categorie 'procedurele activiteiten' een significant verschil aangetroffen. Bij de gecoachte leerkrachten besteedden de leerlingen tijdens de voormeting significant minder tijd aan procedurele activiteiten dan leerlingen bij niet-gecoachte leerkrachten. Verder scoorden gecoachte leerkrachten op de voormeting significant hoger op een van de vijf subschalen van de MIS dan niet-gecoachte leerkrachten, namelijk op de sub-schaal 'gebruik van materiaal en ruimte'. Uit deze resultaten van de toetsing op aanvankelijke verschillen blijkt dat de controle- en de experimentele groep(en) op de voor het onderzoek relevante variabelen nauwelijks van elkaar verschilden. Enkele kleine verschillen werden aangetroffen tussen gecoachte en niet-gecoachte leerkrachten.

Ter beantwoording van de vraagstellingen 1-3 werden eerst vergelijkingen uitgevoerd tussen de onderscheiden groepen voor de scores op de nametingen (nameting én retentiemeting). Vervolgens werden de verschillen getoetst tussen nameting en retentiemeting om te zien of eventuele effecten van de nascholing een duurzaam karakter hadden of niet.

7.1 Effecten van de nascholing

Tabel 2 bevat een overzicht van de gegevens betreffende de tijdsbesteding voor de voor het onderzoek relevante variabelen van het observatie-instrument COMMIT en de beoordelingsschaal MIS. Tabel 3 geeft de resultaten weer van de uitkomsten van de statistische toetsen (uitgedrukt in *t*-waarden). De scores van de nameting en de retentie-meting werden gemiddeld tot één score ter bepaling van het algemeen effect van de nascholing. Getoetst werden de verschillen in deze scores tussen de experimentele groep en de controlegroep en tussen de experimentele groepen onderling, waarbij de scores van de voormeting fungeerden als covariaat. De uitkomsten van deze toet-

singen staan vermeld onder het hoofd 'trainingseffecten'. Bij de interpretatie van deze uitkomsten dienen de gegevens uit Tabel 2 betrokken te worden.

Effecten van het nascholingsprogramma

De gegevens betreffende de vergelijking tussen de experimentele groep (gecoachte en niet-gecoachte leerkrachten) en de controlegroep laten zien dat het nascholingsprogramma een significant effect heeft op het taakgerichte gedrag van de leerlingen. De leerlingen in de experimentele groep zijn meer taakgericht bezig dan de leerlingen in de controlegroep: 80% (of 32.0 minuten: de gemiddelde score van de na-nen retentiemeting) versus 72% (28.7 minuten). Dit verschil is significant ($p < .01$). Leerlingen in de experimentele groep besteden significant minder tijd aan wachten op de leerkracht en aan niet-taakrelevante activiteiten.

In Tabel 3 worden ook de uitkomsten weergegeven met betrekking tot de gemiddelde taakgerichte leertijd in de twee onderscheiden settings: instructie/uitleg en zelfstandige verwerking. In beide settings zijn de leerlingen in de experimentele groep significant taakgerichter dan leerlingen in de controlegroep. Gedurende de uitleg-fase van de les besteden leerlingen in de experimentele groep gemiddeld 86% van de geobserveerde tijd aan hun leertaken (34.4 minuten) en gedurende de zelfstandige verwerkingsfase van de les gemiddeld 78% (31.0 minuten). Voor de leerlingen uit de controlegroep bedragen deze cijfers respectievelijk 77% (30.9 minuten) en 66% (26.6 minuten). Deze resultaten laten zien dat de leerlingen van leerkrachten die aan de nascholing hebben deelgenomen meer taakgericht zijn dan de leerlingen van leerkrachten die niet aan de nascholing hebben deelgenomen.

Gegevens over aspecten van het instructie- en klasmanagementgedrag van leerkrachten zijn eveneens opgenomen in Tabel 3. Significante verschillen zijn aangetroffen tussen de experimentele en de controlegroep voor de categorieën 'terugblik' en 'niet-actief lesgedrag'. De leerkrachten uit de experimentele groep besteden significant meer tijd aan het houden van een terugblik en significant minder tijd aan niet-op-instructiegerichte activiteiten dan de leerkrachten uit de controlegroep (5% versus 3% en 1% versus 6%).

Tot slot bevat Tabel 3 de resultaten van de Management- en Instructie Schaal (MIS). Op vier van de vijf subschalen behalen getrainde leerkrachten significant hogere scores dan niet-getrainde leerkrachten. Alleen voor de subschaal 'gebruik van materiaal en ruimte' is geen significant verschil gevonden. Geconcludeerd kan worden dat de training een positief effect heeft gehad op de instructie- en de klasmanagementvaardigheden van de leerkrachten.

Effecten van coaching

Tussen gecoachte en niet-gecoachte leerkrachten zijn geen significante verschillen aangetroffen. Coaching heeft geen effect gehad op de taakgerichte leertijd van de leerlingen of op het instructie- en klasmanagementgedrag van de aan de nascholing deelnemende leerkrachten.

Effecten van de toepassingstijd

Aangenomen werd dat leerkrachten die vijf jaar geleden aan de nascholing hadden deelgenomen door een langere toepassingstijd hoger zouden scoren dan leerkrachten die twee jaar geleden aan de nascholing hadden deelgenomen. Deze verwachting is niet bewaarheid. Tussen de beide groepen van leerkrachten zijn geen significante verschillen gevonden. Hierbij moet echter worden aangetekend dat de experimentele groep van 1989/90 slechts uit twee leerkrachten bestaat. De uitkomsten van deze vergelijking dienen derhalve met de nodige voorzichtigheid te worden geïnterpreteerd.

7.2 Retentie-effecten

Om de duurzame effecten van het nascholingsprogramma vast te stellen werden de verschillen tussen de retentie-scores en de scores van de nameting (uitgedrukt als verschilcores) van de experimentele groep (gecoachte en niet-gecoachte leerkrachten) en de controlegroep met elkaar vergeleken, eveneens de verschillen tussen de experimentele groepen onderling. Hierbij werden steeds de scores van de voormeting gebruikt als covariaten. De uitkomsten van deze vergelijkingen zijn samengevat in Tabel 3 onder het hoofd 'retentie-effecten'.

Effecten van het nascholingsprogramma op de retentie

Na controle van mogelijke verschillen op de

voormeting, zijn tussen de experimentele groep en de controlegroep geen significante verschillen aangetroffen. Men mag hieruit afleiden dat de scores van de getrainde leerkrachten in de periode tussen de nameting en de retentiemeting, in vergelijking met die van niet-getrainde leerkrachten, niet significant zijn toe- of afgenomen. De resultaten die behaald zijn direct na de training blijven voor de meeste variabelen na verloop van tijd op hetzelfde niveau liggen.

Effecten van coaching op de retentie

Tussen gecoachte en niet-gecoachte leerkrachten zijn geen significante verschillen gevonden. Ondanks de toevoeging van coaching aan de nascholing blijken gecoachte leerkrachten tussen nameting en retentiemeting niet meer vooruit of achteruit te gaan dan niet-gecoachte leerkrachten.

Effecten van de toepassingstijd op de retentie

Drie significante verschillen zijn geconstateerd tussen de getrainde leerkrachten uit 1986/87 en 1989/90. De getrainde leerkrachten uit 1989/90 scoorden lager in de categorie 'wachten op de leerkracht' en hoger in de categorie 'geleide oefening' en op de subschaal 'instructievaardigheden' van de MIS. De gevonden verschillen zijn tegengesteld aan de geformuleerde verwachting. Leerkrachten die de nascholing twee jaar geleden hadden gevolgd scoorden op deze variabelen tussen de nameting en de retentiemeting hoger dan de leerkrachten die de nascholing vijf jaar geleden hadden gevolgd. Gezien het aantal variabelen dat in de toetsing is betrokken, zijn de verschillen tussen beide groepen niet groot.

7.3 Transferpatronen

In Figuur 1 zijn enkele effecten van de training schematisch weergegeven door middel van transferpatronen. Deze transferpatronen maken zichtbaar welke veranderingen tussen voor-, na- en retentiemeting zijn opgetreden. Het transferpatroon met betrekking tot de taakgerichte leertijd van de leerlingen en het geven van een terugblik door de leerkracht kan in termen van Den Ouden (1992) omschreven worden als een 'veranderingscontinueringspatroon' en wel als een 'voortgezet veranderingsproces'. Bij 'geleide oefening' is sprake van

een gestaakt veranderingsproces: direct na de training vindt verandering van gedrag plaats maar deze verandering verdwijnt na verloop van tijd. Er treedt geen transfer op. Bij 'gerichte controle' ook niet. Hier vindt geen implementatie en dus ook geen transfer plaats. Er is geen intentie tot verandering te constateren. De twee weergegeven gedragspatronen voor 'instructievaardigheden' en de 'organisatie van de instructie' zijn indicatief voor de scores op de Management- en Instructie Schaal (MIS). Op alle vijf subschalen is sprake van een 'veranderingscontinueringspatroon'. Alleen bij het 'gebruik van materiaal en ruimte' laat de controlegroep tussen de nameting en de retentiemeting een sterkere stijging zien op deze subschaal van de MIS dan de experimentele groep (figuur hier niet opgenomen; zie Tabel 2).

7.4 Leerprestaties

Het effect van het nascholingsprogramma op de leerprestaties van de leerlingen (vraagstelling 4) is samengevat in Tabel 4. Om verschillen in testlengte en testsoorten voor rekenen en begrijpend lezen te verdisconteren, werden alle ruwe leerlingcores gestandaardiseerd tot z-scores. Verschillen tussen de klassen werden getoetst met behulp van variantie-analyse (eenzijdig).

Bij middeling van de retentie-scores over de klassen heen zijn geen significante verschillen gevonden. De klassen met getrainde leerkrachten scoren niet hoger op de drie schooltoetsen dan de klassen met niet-getrainde leerkrachten. Ook per klas of groep zijn geen systematische verschillen in retentie-scores gevonden tussen getrainde en niet-getrainde leerkrachten (zie Raemaekers & Veenman, 1994). In eerder onderzoek zijn geen verschillen geconstateerd in de scores voor de nameting (Roelofs, Veenman & Voeten, 1994). Verder zijn in deze retentiestudie geen verschillen in leerprestaties gevonden tussen de klassen met gecoachte of niet-gecoachte leerkrachten en tussen de klassen uit de experimentele groep van 1986/87 en de experimentele groep van 1989/90. Bij de laatste vergelijking zijn alleen niet-gecoachte leerkrachten betrokken, omdat coaching in de experimentele groep van 1986/87 niet werd toegepast. Deze resultaten met betrekking tot de leerprestaties van de leerlingen dienen met de nodige voorzichtigheid te worden geïnterpre-

teerd omdat gegevens betreffende leerprestaties voor de aanvang van de training niet voorhanden zijn. Om deze reden heeft geen controle kunnen plaatsvinden op aanvankelijke verschillen in leerprestaties tussen de groepen².

8 Discussie

In deze studie werden de lange-termijn effecten van het nascholingsprogramma onderzocht. Aan de hand van de verzamelde observatiegegevens - twee en vijf jaar na afloop van de nascholing - kon worden vastgesteld dat de getrainde leerkrachten nog steeds gebruik maakten van de in de nascholing geleerde instructie- en klasmanagementvaardigheden. De meeste van de tijdens de nascholing geleerde vaardigheden bleken een duurzame toepassing gevonden te hebben in de klas.

Eén strategie voor transferbevordering bleek echter niet succesvol. Het geven van coaching aan leerkrachten resulteerde niet in duurzame effecten op het taakgerichte gedrag van de leerlingen of op de onderwijsvaardigheden van de leerkrachten. Deze uitkomst kan mogelijk verklaard worden doordat de leerkrachten na afloop van de nascholing geen coaching meer ontvingen. Het ontbreken van schoolbegeleiders aan mogelijkheden voor follow-up coaching. Blijvende ondersteuning op de werkplek kon niet gerealiseerd worden, waardoor de direct na de training gevonden coachingseffecten na verloop van tijd 'verdampen.'

Effecten als gevolg van de toepassingstijd werden nauwelijks gevonden. Waar verschillen werden gevonden waren deze in het voordeel van de groep van getrainde leerkrachten die het nascholingsprogramma twee jaar geleden hadden gevolgd. Een mogelijke verklaring kan zijn dat de leerkrachten uit 1989/90 het belang van de betreffende vaardigheden nog niet vergeten waren. Gezien de omstandigheid dat deze groep slechts twee leerkrachten telde, dient een hypothese omtrent verdere perfectionering van getrainde vaardigheden in de tijd in toekomstig onderzoek nader getoetst te worden.

Vergeleken met de leerkrachten uit de controlegroep boekten de leerkrachten uit de experimentele groep tussen de nameting en de retentiemeting geen verdere vooruitgang met

betrekking tot de uitvoering van de gewenste instructie- en klasmanagementvaardigheden. Uit de resultaten met betrekking tot het interview-instrument 'Gebruiksniveaus van innovatie' (Hall & Loucks, 1977; Hord, Rutherford, Huling-Austin & Hall, 1987) dat in de reëntiestudie eveneens werd gebruikt, bleken de getrainde leerkrachten te functioneren op niveau 3 (mechanisch gebruik) en niveau 4A (routine) (zie Raemaekers & Veenman, 1994). Een stabiel, routinematig patroon van gebruik bleek de werkwijze van de leerkrachten het beste te karakteriseren. Dit niveau van gebruik kan mogelijk verklaard worden door de wijze waarop de instructie- en klasmanagementvaardigheden werden geoperationaliseerd. De meeste vaardigheden waren zeer concreet omschreven. Dergelijke nauw omschreven vaardigheden in gedragstermen bieden mogelijk weinig ruimte voor verdere verfijning of voor een frequenter toepassing. In een les, bijvoorbeeld, is gerichte controle afhankelijk van de problemen die leerlingen hebben met zelfstandig werken. Geen enkele leerkracht zal zijn leerlingen minder zelfstandig laten werken om veel gerichte controle te kunnen toepassen.

Hoewel de getrainde leerkrachten in hun klassen een hogere taakgerichte leertijd realiseerden dan niet-getrainde leerkrachten, leidde dit verschil niet tot een verschil in leerprestaties. Dat meer taakgerichte leertijd niet resulteerde in meer leerwinst kan mogelijk op twee manieren verklaard worden. De eerste verklaring is dat de leerkrachten méér leertijd als een doel op zich beschouwd hebben, wat geleid kan hebben tot meer van hetzelfde. In deze studie zijn geen gegevens verzameld over de wijze waarop de taakgerichte leertijd werd ingevuld. We weten derhalve niet of de geboekte winst aan leertijd besteed werd aan de juiste leertaken. Leertijd werd op een kwantitatieve en niet op een kwalitatieve wijze gemeten. In toekomstige nascholingsprogramma's dienen de leerkrachten meer aandacht te besteden aan de vraag: tijd voor welke leertaken? De kwaliteit van de uitgevoerde leertaak bepaalt wat en hoeveel er geleerd wordt. De tweede verklaring is dat de nascholing het primaat legde bij de instructie van de leerkracht en niet bij het leren van de leerlingen. Leerkrachten zouden meer moeten denken in termen van de gewenste denk- of leerprocessen die ze bij de leerlingen

willen realiseren. Voor toekomstige trainingen betekent dit een verschuiving van denken in instructie-activiteiten naar denken in leeractiviteiten (cf. Simons, 1995). Op deze wijze wordt de invulling van de taakgerichte leertijd direct gekoppeld aan het leren van de leerlingen.

Een van de vaardigheden waarvan het belang voor combinatieklassen wordt benadrukt is het geven van geleide oefening. Twee tot vijf jaar na de training blijkt het effect van de nascholing voor deze vaardigheid verdwenen te zijn (zie Figuur 1). Geleide oefening is belangrijk voor het geven van onderwijs in combinatieklassen omdat via geleide oefening de leerlingen worden voorbereid op het zelfstandig verwerken of toepassen van de leerstof. Pas wanneer de leerkracht zich er van overtuigd heeft dat de leerlingen de leerstof begrepen hebben, kan hij de leerlingen in een groep zelfstandig laten werken en zich wenden tot een andere groep. Uit de gegevens van de voormeting blijkt (zie Tabel 2) dat de leerkrachten weinig gebruik maakten van geleide oefening. Direct na de nascholing werd er significant meer gebruik van gemaakt, maar vergeleken met de voormeting niet overtuigend meer. Twee tot vijf jaar na de nascholing is dit effect verdwenen. Deze terugval in gebruik kan mogelijk verklaard worden doordat de leerkrachten weinig succesvol zijn geweest bij de toepassing van deze vaardigheid in de complexe organisatie van een klas met twee of meer groepen. Het kan ook zijn dat de leerkrachten te weinig ondersteuning uit de werkomgeving hebben ontvangen. In de opzet en uitvoering van de nascholing is de meeste aandacht uitgegaan naar het nemen van transferbevorderende maatregelen vóór en tijdens de nascholing. Aan het traject is weinig aandacht besteed. Zo hebben leerkrachten geen follow-up training ontvangen, is weinig gedaan aan nazorg. In de toekomst zou daarom meer aandacht gegeven dienen te worden aan transferbevorderende organisatiemaatregelen, maatregelen die liggen op het niveau van de school als organisatie (zoals follow-up activiteiten, evaluatie, feedback, peer-coaching).

Noten

- 1 De eerste twee versies van het trainingsprogramma verschenen onder de titel *Omgaan met Combinatieklassen*. Na de twee evaluatiestudies werd het programma herzien en werd het vanaf de derde druk uitgebracht onder de titel *Effectieve instructie en doelmatig klasmanagement* (Veenman, Lem, Roelofs & Nijssen, 1994).
- 2 In de retentiestudie konden de scores voor de schooltoetsen uit de tweede evaluatiestudie niet als covariaten gebruikt worden omdat de meeste leerkrachten twee en vijf jaar na afloop van de nascholing een andere klas hadden met andere leerlingen.

Literatuur

- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research, *Personnel Psychology*, 41(1), 63-105.
- Bennett, B. (1987). *The effectiveness of staff development training practices: A meta-analysis*. Unpublished doctoral dissertation, University of Oregon.
- Broad, M. L., & Newstrom, J. W. (1992). *Transfer of training: Action-packed strategies to ensure high payoff from training investments*. Reading: Addison-Wesley.
- Commissie Evaluatie Basisonderwijs. (1994). *Inhoud en opbrengsten van het basisonderwijs*. Leiden: Distributiecentrum DOP.
- Evertson, C. M., Emmer, E. T., Sanford, J. P., & Clements, B. S. (1983). Improving classroom management: An experiment in elementary school classrooms. *Elementary School Journal*, 84, 173-188.
- Good, T. L., Grouws, D. A., & Ebmeier, H. (1983). *Active mathematics teaching*. New York: Longman.
- Hall, G. E., & Loucks, S. F. (1977). A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal*, 14, 263-276.
- Hord, S. M., Rutherford, W. L., Huling-Austin, L., & Hall, G. E. (1987). *Taking charge of change*. Alexandria: Association for Supervision and Curriculum Development.
- Joyce, B. R., & Showers, B. (1980). Improving inservice training: The messages of research. *Educational Leadership*, 37, 379-385.
- Joyce, B. R., & Showers, B. (1995). *Student achievement through staff development: Fundamentals of school renewal* (2nd ed.). New York: Longman.
- Ouden, M. D. den (1992). *Transfer na bedrijfsopleidingen: Een veldonderzoek naar de rol van voornemens, sociale norm, beheersing en sociale steun bij opleidingstransfer*. Amsterdam: Thesis Publishers.
- Raemaekers, J., & Veenman, S. (1994). *Institutionalisering van teamgerichte nascholing: Duurzame effecten van de cursus Omgaan met Combinatieklassen*. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.
- Roelofs, E., Veenman, S., & Raemaekers, J. (1993). Teamgerichte nascholing en coaching: Effecten op het didactisch handelen van leerkrachten in combinatieklassen. *Pedagogische Studiën*, 70, 352-367.
- Roelofs, E., Veenman, S., & Voeten, M. (1994). Effecten van een teamgerichte nascholing en van coaching op leerprestaties. *Pedagogische Studiën*, 71, 442-456.
- Rosenshine, B. V., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376-391). New York: MacMillan.
- Simons, P. R. J. (1992). De toepassing van transferbevorderende maatregelen in arbeidsorganisaties. *Tijdschrift voor Onderwijsresearch*, 17, 26-37.
- Simons, R. J. (1995). *Leren en werken*. Nijmegen: Vakgroep Onderwijskunde, K.U. Nijmegen.
- Tulder, M. van. (1992). *Nascholing en onderwijsvernieuwing*. (Academisch proefschrift, K.U. Nijmegen) Nijmegen: Universiteitsdrukkerij.
- Tyler, S. (1979). Time-sampling: A matter of convention. *Animal Behavior*, 27, 801-810.
- Veenman S., Lem, P., & Roelofs, E. (1990). Omgaan met combinatieklassen: Een onderzoek naar de effecten van een teamgericht nascholingsprogramma. *Pedagogische Studiën*, 67, 45-58.
- Veenman, S., Lem, P., Roelofs, E., & Nijssen, F. (1994). *Effectieve instructie en doelmatig klasmanagement: Een schoolverbeteringsprogramma voor enkelvoudige en combinatieklassen* (5de druk). Amsterdam: Swets & Zeitlinger.
- Veenman, S., Lem, P., Voeten, M., Winkelmolen, B., & Lassche, H. (1986). *Onderwijs in combinatieklassen*. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs.
- Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

Manuscript aanvaard 22-9-1995

Auteurs

S. Veenman is als universitair hoofddocent werkzaam bij de Vakgroep Onderwijskunde, Katholieke Universiteit Nijmegen.

J. Raemaekers was ten tijde van dit onderzoek verbonden als onderzoeksmedewerker aan de Vakgroep Onderwijskunde, Katholieke Universiteit Nijmegen. Momenteel is hij als schoolbegeleider werkzaam bij de Stichting Schoolpedagogisch Centrum Westelijke Mijnstreek te Sittard.

Adres: Vakgroep Onderwijskunde, Katholieke Universiteit Nijmegen, Postbus 9104, 6500 HE Nijmegen.

Abstract

Retention effects of a staff-development program for teachers in multi-grade classes.

S. Veenman & J. Raemaekers. *Pedagogische Studiën*, 1996, 73, 357-371.

This study describes the long-term effects of a staff-development program based on selected findings

from teaching-effectiveness research in schools with multi-grade or mixed-age classes. The short-term effects of this program were examined in two studies directed at schools with multi-grade classes. The first improvement study was conducted in the school year 1986/87; the second improvement study was conducted in 1989/90. In the latter study, the effects of coaching in addition to participation in the staff-development program were also evaluated. In 1992, a retention or follow-up study was conducted. A quasi-experimental, treatment-control group design was used to test the long-term effects of the program *Dealing with Multi-Grade Classes* and the effects of coaching. Based on pre-, post-training classroom observations, the follow-up study revealed a significant treatment effect for the time-on-task levels of the pupils in the multi-grade classes and for the instructional and classroom management skills of the teachers. No significant differences were found between the coached and uncoached teachers and between the teachers who followed the program either two or five years ago. No significant differences were found between the post-test and the retention test. This suggests that the training results were quite stable. No significant differences in achievement were found between the pupils in classes with trained teachers and the pupils in classes with untrained teachers.

Appendix

Tabel 1

Overzicht van het meetinstrumentarium en het aantal deelnemende leerkrachten aan de eerste en tweede evaluatiestudie en de retentiestudie

Instrument	Eerste studie (1986/87)			Tweede studie (1989/90)			Retentiestudie (1992)		
	Exp. 1 <i>n</i>	Exp. 2 <i>n</i>	Con. <i>n</i>	Exp. 1 <i>n</i>	Exp. 2 <i>n</i>	Con. <i>n</i>	Exp. 1 <i>n</i>	Exp. 2 <i>n</i>	Con. <i>n</i>
COMMIT (observatie)	17	nvt	9	10	18	14	10	8	11
MIS (observatie)	17	nvt	9	10	18	14	10	8	11
Toetsen (leerprestaties)	nvt	nvt	nvt	23	24	16	12	14	14

Opm.: Exp. 1 = experimentele groep met niet-gecoachte leerkrachten; Exp. 2 = experimentele groep met gecochte leerkrachten; Con. = controlegroep

Tabel 2

Gemiddelde tijdsbesteding voor de categorieën van de COMMIT en de subschalen van de MIS voor de voormeting, nameting en retentiemeting

Experimentele groep (n=18)

Controlegroep
(n=11)

Variabelen	Gecoacht (n=8)						Niet-gecoacht (n=10)						Totaal (n=18)						
	1989/90 (n=8)		1986/87 (n=8)		1989/90 (n=2)		1989/90 (n=2)		1989/90 (n=2)		1989/90 (n=2)		1989/90 (n=2)		1989/90 (n=2)		1989/90 (n=2)		
	Voor	Na	Ret	Voor	Na	Ret	Voor	Na	Ret	Voor	Na	Ret	Voor	Na	Ret	Voor	Na	Ret	
COMMIT																			
Leerlinggedrag (min.)																			
Taakgericht	28.1	30.8	32.6	28.0	33.0	31.3	29.4	31.2	33.9	28.3	32.7	31.8	28.2	31.9	32.1	28.7	27.3	30.1	
Procedurale activiteiten	2.8	2.3	2.5	5.0	3.3	2.8	2.7	2.2	2.3	4.5	3.1	2.7	3.8	2.7	2.6	3.4	3.7	2.7	
Wachten op de leerkracht	2.1	1.3	0.8	1.0	0.4	0.9	1.7	1.5	0.3	1.1	0.6	0.8	1.6	0.9	1.4	1.4	1.4	1.1	
Niet met de opdracht bezig	7.0	5.6	4.2	6.0	3.3	5.1	6.3	5.1	3.5	6.1	3.7	4.8	6.5	4.5	4.5	6.6	7.5	6.1	
Setting en taakgericht (%)																			
Instructie en taakgericht	71.8	81.6	86.0	76.7	86.6	89.4	77.2	82.3	90.1	76.8	85.7	89.6	74.6	83.9	88.0	75.4	72.7	81.8	
Verwerking en taakgericht	69.6	75.4	77.9	68.3	81.4	75.6	70.8	76.0	78.8	68.8	80.3	76.2	69.2	78.1	77.0	69.6	63.5	69.5	
Leerkrachtgedrag (min.)																			
Instructie:																			
- Terugblik	0.1	2.2	3.3	0.5	1.9	0.6	0.0	2.8	2.5	0.4	2.1	1.0	0.3	2.1	2.0	0.2	0.4	0.4	
- Uitleg	20.0	18.8	20.9	9.1	9.0	18.1	22.8	22.1	30.2	11.8	11.6	20.5	15.5	14.8	20.7	20.3	19.1	24.6	
- Geleide oefening	0.9	1.0	0.0	2.9	6.3	0.4	0.0	0.8	1.6	2.3	5.2	0.6	1.7	3.3	0.4	2.0	0.9	0.9	
Begeleiding bij verwerking:																			
- Individuele hulp	6.1	6.9	6.5	14.7	12.4	11.8	6.9	3.4	3.4	13.1	10.6	10.2	10.0	9.0	8.5	7.4	6.9	5.7	
- Gerichte controle	5.8	4.6	4.3	4.0	3.9	3.4	5.1	5.7	0.9	4.2	4.3	2.9	4.9	4.4	3.5	3.2	4.0	3.0	
Organisatie:																			
- Organisatie/leswisselingen	5.8	5.6	4.0	6.0	5.4	3.6	4.3	3.8	1.3	5.6	5.1	3.1	5.7	5.3	3.5	4.6	5.5	3.6	
- Niet-actief lesgedrag	1.3	1.0	1.2	3.0	1.1	2.2	1.0	1.5	0.2	2.6	1.2	1.8	2.0	1.1	1.5	2.2	3.3	1.8	
Subschalen MIS																			
(gesommeerde itemscores)																			
Instructievaardigheden	22.4	25.1	27.5	24.3	29.0	25.4	21.0	25.0	30.3	23.7	28.2	26.4	23.1	26.8	26.9	22.8	22.0	23.6	
Organisatie vd instructie	17.5	20.4	22.4	16.8	20.7	20.9	18.4	20.8	22.5	17.1	20.7	21.2	17.3	20.6	21.7	17.8	18.0	18.7	
Gebruik materiaal/ruimte	24.6	27.5	28.8	21.1	25.2	26.4	25.5	26.3	29.5	22.0	25.4	27.0	23.2	26.3	27.8	24.0	24.6	27.8	
Afstemming vd instructie	21.4	24.1	26.3	20.7	24.4	24.8	22.0	24.3	26.5	21.0	24.4	25.2	21.2	24.3	25.7	21.5	21.3	23.4	
Omgaan met storingen	20.4	24.2	26.3	19.7	24.6	23.1	20.1	24.6	27.5	19.8	24.6	24.0	20.1	24.4	25.0	19.8	20.2	22.0	

Opm.: Lesduur = 40 minuten. Voor = voormeting, Na = nameting, Ret = retentiemeting

Tabel 3

Training- en retentie-effecten: Resultaten van de t-toetsen voor de scores van de observatiecategorïeën van de COMMIT en van de subschalen van de Management en Instructie Schaal (MIS) van de retentie-studie (met de score van de voormeting als covariaat)

Observatiecategorïeën	Trainingseffecten			Retentie-effecten		
	Training: exp. vs. con.	Coaching: gecoacht vs niet- gecoacht	Tijd: 1986/87 vs. 1989/90	Training: exp. vs.con.	Coaching: gecoacht vs.niet- gecoacht	Tijd: 1986/87 vs. 1989/90
	t	t	t	t	t	t
COMMIT						
<i>Leerlinggedrag</i>						
Taakgericht	3.83**	-0.28	-0.19	-1.29	0.69	-1.60
Procedurale activiteiten	-1.69	0.24	-0.10	1.15	0.38	-0.32
Wachten op de leerkracht	-2.19*	-0.20	0.12	-0.11	-0.31	2.08*
Niet met de opdracht bezig	-3.14**	0.26	-0.01	1.07	-1.19	1.84
<i>Setting en taakgerichtheid</i>						
Instructie en taakgericht (%)	3.44**	-0.36	0.41	-0.80	-0.14	-0.52
Verwerking en taakgericht (%)	3.86**	-0.37	0.39	-1.97	1.02	-1.44
<i>Leerkrachtgedrag</i>						
Terugblik	4.93**	1.78	-2.05	-0.26	1.65	-0.60
Geleide oefening	1.56	-1.75	0.68	-1.47	0.91	-2.56*
Gerichte controle	-0.25	0.54	0.43	-0.45	0.96	1.22
Organisatie/leswisseling	-1.52	1.11	0.89	-0.09	0.37	0.27
Niet-actief lesgedrag	-2.59*	0.27	-0.09	1.52	0.29	1.28
<i>Subschalen MIS</i>						
Instructievaardigheden ($\alpha = .90$)	4.89**	-0.88	-0.87	-0.27	1.04	-3.75**
Organiseren vd instructie ($\alpha = .84$)	4.10**	0.22	-0.03	0.56	0.73	-0.71
Gebruik materiaal/ruimte ($\alpha = .66$)	1.95	1.00	-0.24	-0.83	-0.46	-0.62
Afstemming vd instructie ($\alpha = .86$)	3.37**	0.24	-0.22	-0.48	0.69	-0.91
Omgaan met storingen ($\alpha = .91$)	3.12**	0.09	-0.90	-0.51	0.78	-1.62

Opm.: Exp. = experimentele groep; Con. = controlegroep. Tijd = toepassingstijd.

De leerkrachtgedragingen uitleg en individuele hulp vormden qua tijdsbesteding geen oogmerk van de cursus en zijn derhalve niet in de toetsing opgenomen.

* = $p < .05$; ** = $p < .01$

Tabel 4

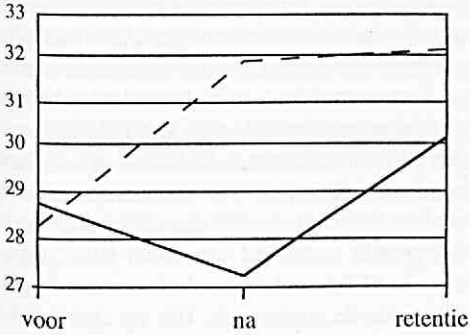
ANOVA-testresultaten voor de z-scores betreffende de leerprestaties (gemiddeld over de groepen drie tot en met acht)

Test	Experimentele groep			Controlegroep			F	p
	n	Gem	SD	n	Gem	SD		
Technisch lezen	630	0.01	1.00	349	-0.02	0.99	0.21	.64
Begrijpend lezen	634	0.04	1.01	329	-0.07	0.97	2.63	.11
Rekenen	632	-0.02	1.00	337	0.04	1.00	0.90	.34

Opm.: Experimentele groep $n = 26$ klassen, controlegroep $n = 14$ klassen.

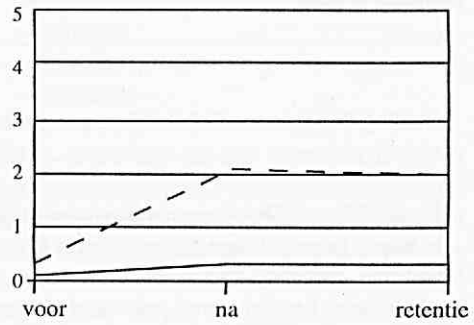
Taakgerichte leertijd

(lesduur 40 minuten)



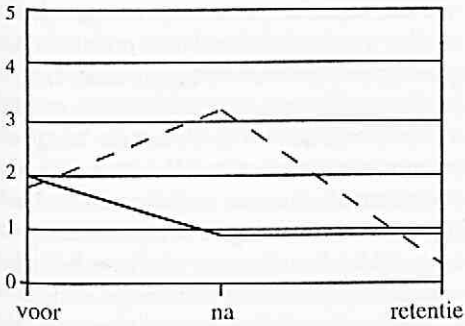
Terugblik

(lesduur 40 minuten)



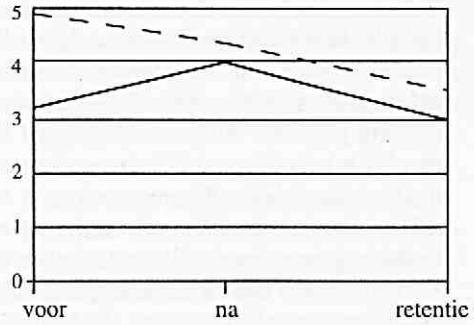
Geleide oefening

(lesduur 40 minuten)



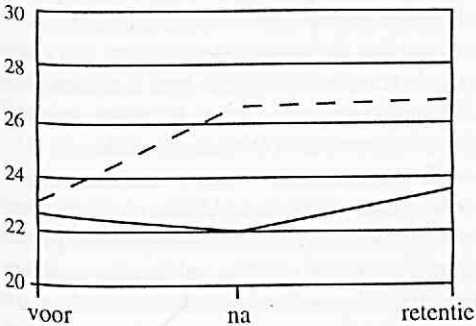
Gerichte controle

(lesduur 40 minuten)



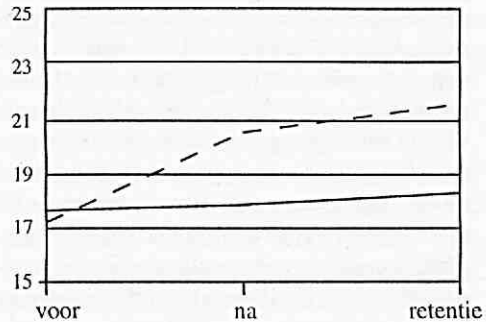
Instructievaardigheden (MIS)

(min.= 7, max.= 35)



Organisatie van de instructie (MIS)

(min.= 5, max.= 25)



--- Experimentele groep — Controlegroep

Figuur 1. Enkele voorbeelden van transferpatronen

Het AERA-Congres 1996, New York

Inleiding (N. Verloop, ICLON, Rijksuniversiteit Leiden)

Het jaarlijkse AERA-congres is zonder twijfel de meest grootschalige happening op het gebied van onderwijsonderzoek ter wereld. Al enkele jaren ligt het aantal presentaties tussen de vier- en de vijfduizend. Het congres dat van 8 tot 12 april in New York werd gehouden vormde hierop, uiteraard, geen uitzondering. Opvallend is dat de Nederlandse onderwijsonderzoekers daarbij steeds sterk vertegenwoordigd zijn. Ik heb geen analyse gemaakt van het relatieve aandeel van de Nederlandse onderzoekers, maar wijs bijvoorbeeld op een analyse die vorig jaar door White in *Educational Researcher* is gepubliceerd. Hij ging onder meer na uit welke landen de presentaties op het AERA-congres van 1994 afkomstig waren. Omdat er geen reden is om aan te nemen dat deze getallen van jaar tot jaar ingrijpend zullen verschillen, vermeld ik hier kort zijn bevindingen. In totaal werden door auteurs uit 46 landen presentaties verzorgd. Het overgrote deel was vanzelfsprekend afkomstig uit de VS. In de totale lijst van 46 landen stond Nederland (met 37 presentaties) op de vijfde plaats, na de VS, Canada, Australië en Engeland. Met andere woorden, van alle niet-Engelssprekende landen scoorde Nederland het hoogste. Als we bovendien in aanmerking nemen dat het afwijzingspercentage voor veel divisies inmiddels al rond de 40% ligt en dat voorstellen van buiten de VS niet bepaald extra welwillend worden bejegend, kan dit als een opvallende prestatie gelden. Het geeft in elk geval weinig grond voor het pessimisme omtrent de internationale oriëntatie van de Nederlandse onderwijskunde, waarmee sommige beschouwers van het Nederlandse onderwijsonderzoek de laatste tijd behept lijken te zijn.

Het voortdurend speuren naar nieuwe formats die afwijken van de traditionele symposium-, paper- en posterpresentaties had ook

deze keer weer nieuwe varianten opgeleverd. Inmiddels komt men in het programmaboek zo'n 20 formats tegen, waaronder 'book sessions' (waarin een nieuwe publicatie wordt gepresenteerd en becommentarieerd), 'consultation sessions' (waarin de mogelijkheid voor het verkrijgen van hulp en advies gegeven wordt) en 'demonstrations'. Niet alle varianten blijken overigens erg zinvol te zijn. Dat gold wel voor een initiatief waarmee dit jaar voor het eerst werd geëxperimenteerd, nl. het als zodanig aankondigen van sessies die sterk op de praktijk gericht waren. Al een aantal jaren stimuleert de AERA praktiserende docenten tot deelname aan de conferentie. Dit, op zich zonder twijfel zinnige, streven tot vergroting van de functionaliteit van het onderwijsonderzoek leidt er soms toe dat onderzoekers die geïnteresseerd zijn in de onderzoeksmatige aanpak van een bepaalde thematiek, verzeild raken in sessies waarin uitsluitend over praktische toepassingsproblemen wordt gesproken. Dit jaar heeft men getracht door middel van speciale symbolen die sessies te labelen die 'of special interest to teachers' waren. Het systeem werkte weliswaar nog lang niet perfect, maar het biedt, bij striktere toepassing in de toekomst, wel een mogelijkheid om het type sessies te bezoeken waarin men geïnteresseerd is.

Het conferentiethema van dit jaar was "Research for education in a democratic society". Het kiezen van een thema voor een conferentie als deze is volstrekt zinloos. Omdat de divisies vrij zijn hun eigen beleid te voeren heeft een dergelijk thema geen enkele invloed op de voorstellen die aangeboden worden. Het enige wat de congresleiding kan doen is achteraf een label plakken op de sessies die 'theme-related' zijn, hetgeen dan ook braaf gebeurde.

Tot slot vermeld ik nog een elk jaar terugkerend hoogtepunt van de AERA, nl. de tentoonstelling waarop werkelijk alle grote uitgeverij hun, veelal zeer recente, publicaties presenteren en verkopen. Dat laatste leidt er steevast toe dat tal van Europese congresgangers met aanzienlijk zwaardere koffers huiswaarts keren dan waarmee zij gekomen zijn.

In de hier volgende rapportage is gekozen voor een divisie-gewijze aanpak in plaats van een aanpak waarbij over allerlei specifieke onderwerpen wordt geschreven. De bedoeling is dat

hierdoor een globaal beeld van het congres als geheel ontstaat, al blijft er natuurlijk, op basis van de interesses van de betrokkenen, sprake van een selectie van datgene wat binnen de diverse divisies aan de orde is geweest. Aan de auteurs is gevraagd vooral in te gaan op algemene trends die binnen hun divisie zichtbaar waren. Van de 11 AERA-divisies komen er 7 aan bod. Zij bestrijken het overgrote deel van hetgeen in het Nederlandse taalgebied onder 'onderwijskunde' of 'onderwijswetenschappen' wordt verstaan.

Literatuur

White, R. (1995). International scholarship and the AERA. *Educational Researcher*, 24(6), 19-21.

Divisie B: Curriculum Studies (M. Mulder, Universiteit Twente)

Ook dit jaar sprong de veelheid aan bijdragen binnen de Divisie Curriculum weer in het oog. In totaal waren er zo'n tachtig sessies bestaande uit verschillende presentaties, alsmede een ledenvergadering en een bestuursvergadering. Er waren 270 voorstellen voor de deadline ingediend. Als de voorstellen die na de deadline bij de organisatiecommissie zijn binnengekomen worden meegeteld, dan ligt dat aantal zelfs boven de 300. Uitsluitend de voorstellen die bijtijds zijn ontvangen zijn beoordeeld. Daarvan waren er 167 ingediend als traditionele paperpresentaties, er waren 41 voorstellen voor symposia en 62 voor experimentele sessies. Het beleid van de AERA is erop gericht het indienen van voorstellen voor de experimentele, meer interactieve sessies te bevorderen. Gezien het aantal van dat soort ingediende voorstellen kan worden opgemerkt dat dit beleid in de praktijk ook redelijk wordt gerealiseerd. Interessant is te vermelden dat de acceptatieratio dit jaar lag op 58%. Dat komt overeen met 157 voorstellen van voldoende, goede en uitstekende kwaliteit. Er waren echter slechts 52 slots in het programma, maar door te onderhandelen met andere Divisies en Special Interest Groups, konden toch alle goedgekeurde voorstellen worden opgenomen in het programma.

Wat de inhoud van het programma betreft, kan worden opgemerkt dat er vier subthema's

waren:

- curriculumonderzoek in de context van de classesituatie;
- curriculumtheorievorming;
- curriculumontwerp, -implementatie en -evaluatie;
- curriculumgeschiedenis, -bewegingen en -praktijken.

Het is ondoenlijk om een overzicht te geven van alle bijeenkomsten binnen de divisie, maar enkele titels van sessies zullen duidelijk maken dat de koppeling tussen de vier genoemde subthema's en de inhoud van de sessies niet altijd duidelijk is.

- Het doordenken van de levens van vrouwelijke leerkrachten: poststructuralistische en materialistische feministische benaderingen.
- Reflecties over de formuleringen van Dewey over democratie en onderwijs: spel als een bron van gemeenschap voor kinderen.
- Vrouwen en curriculum: een feministische exploratie.
- Het herstructureren van onderwijsonderzoek in een problematische democratie.
- Een nieuwe generatie onderzoekers: exploraties in lesbische, homo- en bisexuele problemen in het onderwijs.
- Voor wie en hoe is de vakinhoud van belang: Afrikaans-Amerikaanse leerlingen en het schoolcurriculum.

Deze voorbeelden geven tegelijkertijd aan dat er binnen de Divisie Curriculum grote belangstelling is voor actuele sociaal-maatschappelijke en wetenschapstheoretische problemen en voor de doordenking van de consequenties van die problemen voor de inhoud van het onderwijs voor kinderen en jongeren.

Binnen de divisie is in verschillende andere bijeenkomsten ook aandacht geschonken aan het algemene thema van de conferentie: "Onderzoek voor onderwijs in een democratische samenleving". Zo stond de door de Divisie Curriculum geïnitieerde plenaire bijeenkomst in het teken van verbeelding (imagination), mogelijkheid (possibility) en curriculum, een tamelijk abstracte aanduiding voor een thematiek waarin de rol van de inhoud van het onderwijs voor de vorming van kinderen en jongeren binnen een samenleving, waarvan het democratisch gehalte onder druk staat, centraal

wordt gesteld. Twee eminenties op het terrein van de onderwijskunde, Greene en Sarason presenteerden hun zienswijze op deze problematiek. Beide sprekers gingen in op het belang van onderwijs in de 'arts' (zoals schilderen, muziek, drama en dichtkunst) voor de ontwikkeling van de verbeelding van kinderen en jongeren. Greene betoogde dat deze ontwikkeling van cruciaal belang is. Veel kinderen en jongeren voelen zich nutteloos, wanhopig en zijn het slachtoffer van haat, aldus Greene. Kunstonderwijs kan bijdragen aan de transformatie van hun perceptie van de werkelijkheid en kan een mogelijkheid bieden zich een andere werkelijkheid voor te stellen en aan de realisatie daarvan te werken. Het kunstonderwijs kan aldus kinderen en jongeren een perspectief bieden van waaruit zij betekenis in plaats van betekenisloosheid zoeken, en apathie en geestelijke anesthesie bestrijden. Deze zienswijze gaat terug op de filosofie van Dewey en moet volgens haar niet worden opgevat als het creëren van luchtkastelen: het kunstonderwijs moet direct gerelateerd worden aan de realiteit. Sarason onderstreepte het belang van deze zienswijze en was het op opvallend weinig punten oneens met Greene, ofschoon hij in de herziene versie van zijn 25 jaar oude boek 'Revisiting the Culture of the School and the Problem of Change' weinig woorden wijdt aan het kunstonderwijs en de bijdrage daarvan aan de ontwikkeling van de verbeelding en de mogelijkheid om betekenis te zoeken, bewondering te wekken en de bestaande situatie te transfigureren naar een meer ideale situatie. Sarason wees wel op het belang van de wijze waarop het kunstonderwijs in de praktijk wordt vormgegeven. Hij onderscheidt daarbij productieve en non-productieve onderwijscontexten. De non-productieve vormen voegen weinig toe aan de ontwikkeling van de verbeelding maar komen helaas wel het meest voor in het Amerikaanse onderwijs. Helaas - aldus Sarason - is dit de afgelopen jaren niet verbeterd, vooral ook omdat het onderwijs alle kenmerken mist van zelfcorrigerende en zelfverbeterende systemen die naar kwaliteit streven. En helaas hebben de betrokkenen bij de vormgeving van het onderwijs nog niet het voorstellingsvermogen om het onderwijssysteem te zien als een zelfcorrigerend systeem.

Ofschoon er een kern van waarheid zit in het betoog van Sarason, is de constatering mis-

schien niet geheel juist. Want in een andere, ook zeer druk bezochte bijeenkomst werd gesproken over KEYS, een grootschalig project waarin scholen gebruik kunnen maken van een gestandaardiseerde methodiek voor kwaliteitsbevordering van het onderwijs. Er wordt gebruik gemaakt van een diagnostisch instrumentarium, een kwaliteitscommissie in de school zorgt voor de uitvoering van deze diagnose, de National Educational Association (NEA) verwerkt de gegevens en rapporteert aan de school aan de hand van 39 indicatoren. De presentaties in deze bijeenkomst werden verzorgd door Schneider van NEA, Fullan van de Universiteit van Toronto, Lieberman van Teachers College, Columbia University en Rolles, eveneens van NEA. Professioneel gebruik van een systeem als KEYS kan bijdragen aan de zelfcorrectie en -verbetering waar Sarason reeds jaren voor pleit. De titel van deze bijeenkomst luidde: 'Het gebruik van leerkrachtenonderzoek om lerende organisaties te creëren en de onderwijs-professie te bevorderen'. Dit thema sluit aan bij het algemene beleid van de AERA om onderzoek dat is uitgevoerd door of samen met leraren ook op te nemen in het AERA-programma. Deze trend is overigens al een aantal jaren geleden ingezet. Er is al een aantal jaren plaats voor het presenteren van onderzoek dat is uitgevoerd door of met leraren, maar dit jaar is geprobeerd de leraren nog meer bij de conferentie te betrekken dan in voorgaande jaren. Was dit in het verleden binnen de Divisie Curriculum voornamelijk beperkt tot de rol van de leraar in het implementeren van curriculummaterialen en werd de leraar beschouwd als een lerende, en als consument van onderzoek, en lag de nadruk op implementatie- en 'utility'-onderzoek, nu heeft de AERA-organisatie ook gekozen voor een afzonderlijke dag waarop speciale aandacht is gegeven aan de praktijk. Deze dag heeft een speciale naam meegekregen: 'Sessies met de scholen'.

Ofschoon dit streven met sympathie moet worden bejegend, leidt het wel tot een combinatie van verschillende soorten onderzoek. De reactie van wetenschappelijk onderzoekers is dan al gauw: de kwaliteit van het onderzoek gaat hard achteruit. Er werd bijvoorbeeld in één van de sessies een presentatie gehouden over een nascholingsprogramma op het gebied van archeologie. Docenten van verschillende disci-

plines en onderwijsniveaus konden participeren in het programma. Het programma bestond uit een periode van veldwerk van drie weken in een woestijnachtige omgeving. Het veldwerk bestond onder andere uit het verrichten van opgravingen. Er was een eenvoudig kamp waar de deelnemers konden verblijven. In de presentatie die werd verzorgd door de organisator van deze nascholingsactiviteit, werd uiteengezet welke activiteiten de deelnemers verrichtten en op welke wijze de deelnemers gebruik konden maken van de activiteiten in hun onderwijs. De deelnemers zijn na afloop van het nascholingsprogramma een tijd gevolgd om na te gaan op welke wijze zij gebruik gemaakt hebben van hetgeen zij in het nascholingsprogramma hebben geleerd. Een dergelijke presentatie kan interessant zijn voor docenten en nascholers in dit vakgebied, maar voor wetenschappelijk onderzoekers is het niet veel meer dan een praktijkverhaal zonder wetenschappelijk kader, en kan dan ook moeilijk worden opgevat als een presentatie over wetenschappelijk onderzoek. Misschien moeten de bezoekers van toekomstige AERA's zich meer en meer realiseren dat 'educational research' binnen AERA-verband een rekbaar begrip is waarin uitdrukkelijk plaats is voor praktisch situationeel bepaald onderzoek dat wordt verricht door leraren. Als de AERA als beleid heeft leraren doelbewust bij de conferentie te betrekken, dan is het begrijpelijk dat dit soort onderzoek ook een plaats krijgt binnen het programma. Het is niet ondenkbaar dat leraren daar het meest aan hebben. Maar er ontstaat dan wel een mix van wetenschappelijk onderzoek voor onderzoekers en ideosyncratisch onderzoek voor nascholingsdoeleinden. De toekomst zal uitwijzen of deze combinatie vruchtbaar is. In ieder geval is het bezoekersaantal van de AERA voldoende groot om iedereen een plaats te bieden.

Een interessant fenomeen in dit geheel is dat er dit jaar voor het eerst ook sessies binnen scholen konden worden bezocht. Voor deze bijeenkomsten moest men zich van tevoren opgeven en een eindje reizen, maar het was voor sommigen wel een mogelijkheid een school in New York van binnen te zien.

Zoals bijna ieder jaar was er binnen de Divisie Curriculum weer aandacht voor de vergelijking van de ontwikkelingen in de curriculumtheorie en het curriculumonderzoek in de

Verenigde Staten en Europa. De titel van de eveneens druk bezochte sessie over dit onderwerp was veelbelovend: 'Curriculumtheorie en -onderzoek revisited na 20 jaar. Ligt Europa nog steeds aan kop?' In 1976 is er een bijeenkomst binnen de Divisie Curriculum geweest over hetzelfde onderwerp, waarin is geconcludeerd dat Europa verder was in de curriculumtheorievorming en het curriculumonderzoek dan de Verenigde Staten. De presentaties in deze bijeenkomst werden verzorgd door Hamilton, Weiner, Sander en Calgren (met Kallós). Discussianten waren Kliebard en Yates.

Vanuit de Britse bijdragen werd duidelijk dat inmiddels anders wordt gedacht over talrijke onderwijsproblemen. Veel van de openbare discussies zijn gericht op meer technische aspecten in het onderwijsbestel en op het bevorderen van de effectiviteit van scholen. Volgens de sprekers is dat een beperkte discussie en moet er veel meer aandacht zijn voor de maatschappelijke problemen en de relatie daarvan met het curriculum.

De bijdrage uit Duitsland ging over het falen van het twintig jaar geleden geïntroduceerde nieuwe paradigma voor het onderwijs. Dit paradigma bestond uit een nieuwe visie op de Duitse didactiek, waarbij nadruk kwam te liggen op open communicatie, open curricula en open curriculumontwikkeling. In tegenstelling tot wat werd aangekondigd als een renaissance van de Duitse didactiek moet volgens Sander nu worden geconcludeerd dat de Duitse didactiek dood is en dat zelfs wederopstanding onmogelijk is. Als voornaamste oorzaak voor dit falen wordt aangevoerd dat leerlingen uiteindelijk toch geen macht hebben over hun situatie in het onderwijs en velen zich daarom ook afvragen waarom zij zich druk zouden maken over de democratische discussie. Ook de unificatie en de economische recessie heeft geen goed gedaan aan de verspreiding van de ideeën over de Duitse didactiek. In het Duitse onderwijsbestel (zowel in west als oost) is veel verzet tegen ideologisch getint onderwijs.

Carlgrén gaf een schets van de disconnectie van het onderwijsonderzoek van het onderwijsbeleid en de onderwijspolitiek, respectievelijk de 'National Board of Education' en het Ministerie van Onderwijs in Zweden. De positie van het onderwijsonderzoek is ernstig ver-

slechterd nu het geen integraal deel meer uitmaakt van het onderwijsbeleid en de onderwijspolitiek. Het onderwijsbestel is meer gedecentraliseerd en het onderwijsonderzoek is momenteel op zoek naar nieuwe doelgroepen. Opvallend is dat de decentralisatie heeft geleid tot centrale besturing van het onderwijs op decentraal niveau en dat er sprake is van een tegenstelling tussen meer instrumentele en meer kritische benaderingen op decentraal niveau. De uitdaging voor onderzoekers is volgens Carlgren om het overleg op decentraal niveau tussen de verschillende categorieën betrokkenen beter te structureren.

Nadat de discussianten vanuit hun perspectieven wat grote lijnen hadden getrokken kwam de prangende vraag in de titel van deze sessie uiteindelijk uit de zaal. Wat was nu het antwoord: ligt Europa nog steeds voor? Na enkele ogenblikken doodse stilte sprak Kallós het verlossende woord: 'Ja'. Helaas gaf hij vervolgens geen argumenten voor zijn antwoord. Gelukkig werd in de zaal wat genuanceerder gedacht over dit antwoord. Het antwoord moet natuurlijk zijn: 'dat hangt er van af'. Het hangt af van het onderwerp binnen de curriculumtheorie en het curriculumonderzoek en uiteraard ook van het perspectief van waaruit de vraag wordt beantwoord. Bovendien waren slechts drie landen uit Europa vertegenwoordigd. En waren er geen presentaties waarin de Europese en Amerikaanse ontwikkelingen systematisch werden vergeleken. Dat zou mijns inziens het antwoord op deze vraag moeten zijn. Als men dat wil weten, moet men dat stelselmatig gaan vergelijken en zijn oordeel niet baseren op onafhankelijke analyses van de ontwikkelingen in drie landen.

Divisie C: Learning and Instruction

C - I: Leren en instructie (L. Verschaffel en E. De Corte, Katholieke Universiteit Leuven)

In Divisie C werd heel veel aandacht besteed aan de thema's die ook gedurende de voorgaande jaren het onderzoek rond probleemoplossen en leren sterk in de belangstelling stonden, m.n. de rol van metacognitie en zelfregulatie in (leren) denken en probleemoplossen, de interactie tussen cognitieve en motivationele aspecten, de waarde en haalbaarheid van

de constructivistische visie op leren en onderwijzen, de waarde van het 'situated cognition' paradigma en de diverse interpretaties ervan, het ontwerpen van krachtige leeromgevingen. Onze globale indruk is dat er ten aanzien van deze thema's geen noemenswaardige nieuwe ontwikkelingen te melden vallen. Wel vond er een aantal boeiende confrontaties en discussies plaats en werden tal van empirische gegevens gerapporteerd die de reflectie over en het onderzoek van deze onderwerpen verder vooruit helpen. Omdat we zelf vooral sessies gevolgd hebben waarin vragen omtrent leren en instructie toegespitst werden op het vakgebied 'wiskunde', zullen we ons verslag ook daarop concentreren. Doch vooraf gaan we - exemplarisch - in op enkele sessies waarin de bovengenoemde topics vanuit een meer algemeen perspectief werden behandeld.

In het druk bijgewoond symposium 'Conceptions of learning and understanding', waarin o.a. ook bijdragen van bekende Europese onderwijspsychologen als Entwistle, Säljö en Marton werden gepresenteerd, hield Perkins een opmerkelijke lezing waarin hij zich sterk afzette tegen de klassieke 'representational view of understanding', die uitgaat van de assumptie dat 'understanding' gebaseerd is op de constructie en het gebruik van kennis in de vorm van schematische of propositionele netwerken, regelsystemen e.d. Vertrekend vanuit de idee dat het hebben van begrip van of inzicht in iets enkel maar kan blijken uit 'flexible performance', trachtte Perkins de toehoorders ervan te overtuigen dat het beschikken over en het gebruik maken van (bewuste) representaties noch een noodzakelijke noch een voldoende voorwaarde is om te kunnen spreken van 'understanding'. Volgens hem dient de 'representational view of understanding' vervangen te worden door een 'performance view of understanding'; datgene wat geconstrueerd wordt wanneer er kennis en begrip van iets wordt opgebouwd moet opgevat worden in termen van 'flexible performance capacities'. In hetzelfde symposium lichtte Entwistle zijn ideeën toe over 'knowledge objects' als wezenlijke bestanddelen van de menselijke cognitie en werd door Marton gewezen op de epistemologische moeilijkheden die veroorzaakt worden door het individu van zijn omgeving los te koppelen; deze loskoppeling genereert een

kloof die logisch niet meer overbrugbaar is. In zijn discussiebijdrage reageerde Salomon (fel) op Perkins' betoog; gebruikmakend van voorbeelden zowel uit het dagelijks leven als uit de recente onderzoeksliteratuur, pleitte hij voor het behoud van het representatiebegrip dat door Perkins zo scherp onder vuur was genomen. Zo argumenteerde hij dat het gelijkstellen van 'iets begrijpen' met 'begrijpend gedrag vertonen' en het in vraag stellen van de rol van mentale representaties bij het begrijpen van de werkelijkheid, ons wel zeer (gevaarlijk) dicht brengt bij een behavioristisch georiënteerd gedachtegoed.

In zijn inleiding tot het symposium met de veelbelovende titel 'Situated cognition versus cognitivism: a grounded debate' stelde diSessa de vraag wat de voor- en nadelen zijn van het ijveren voor een paradigma-wisseling binnen een wetenschapsgebied en van het beklemtonen van verschillen tussen de concurrerende paradigma's. Sherin stelde een empirisch onderzoek voor rond (leren) oplossen van fysica-problemen dat als prototypisch voor het traditionele cognitivistische paradigma kon gelden en pleitte op grond daarvan voor het behoud van dit paradigma. Daarna was het de beurt aan Greeno. Eerst zette hij de - inmiddels bekende - verschillen tussen het 'cognitivistische' en het 'situationele' perspectief nogmaals in algemene termen op een rijtje. Daarna deed hij een - niet erg geslaagde - poging om te illustreren hoe deze verschillen (duidelijk) tot uiting kwamen in de wijze waarop (kwalitatieve) onderzoeksgegevens in beide onderzoeksparadigma's geanalyseerd en geïnterpreteerd worden. Tenslotte ging diSessa zelf diepgaand in op de problemen die zich stellen bij de aanvang van een individueel interview. Hoewel zijn bevindingen en interpretaties op zich erg interessant waren, was het voor de (meeste) toehoorders niet zo duidelijk waarin de bijdrage van deze lezing tot het 'grounded debate' over het cognitivistische en het situationele paradigma precies gelegen was.

Mede door het voorkomen van meer radicale en meer gematigde standpunten en meer cognitieve en meer sociale varianten van het constructivisme, blijft de constructivistische visie op leren nog steeds het voorwerp van veel discussie. Dit kwam in talloze sessies tot uiting. Nogal wat sprekers verwezen in hun bijdrage

naar het recente maar inmiddels bekende artikel van Phillips van Stanford University 'The good, the bad, and the ugly: The many faces of constructivism' (Phillips, 1995). Het 'ugly aspect', zegt Phillips, is dat het constructivisme in vele gevallen verworpen is tot een soort religie, tot een ideologie. Er is ook een slechte kant ('bad') aan, namelijk het feit dat er een tendens in zit naar absoluut relativisme. Tenslotte is er evenwel een goede kant ('good'), m.n. de klemtoon op de noodzaak van actieve participatie van de lerende en de erkenning van het sociaal karakter van leren.

In een zeer groot aantal sessies stond de rol van metacognitieve kennis en van zelfregulatie in het (leren) probleemoplossen en het (leren) leren centraal. Daarbij ging opvallend veel aandacht naar de mate waarin de metacognitieve denkbeelden en het zelfregulerend vermogen van subjecten (leerlingen, leerkrachten...) te beïnvloeden is. Dit was o.m. het geval in een paneldiscussie getiteld 'Self-regulation in education: future directions in research and applications', waaraan tal van vooraanstaande onderzoekers in dit gebied deelnamen zoals Schunk, Corno, Pintrich, Pressley, Winne en Zimmerman (Weinstein was afwezig). Naast de vraag hoe de verworven kennis gebruikt kan worden voor het verbeteren van het onderwijzen en leren van zelfregulatievaardigheden, kwamen in deze paneldiscussie ook de volgende thema's aan bod: welke zijn de 'component processes' van zelfregulatie, welke zijn de prioritaire vragen en problemen voor het onderzoek in de nabije toekomst, en welke vernieuwingen en verbeteringen in de onderzoeksmethodologie zijn er nodig? Interessant was ook het door Tobias georganiseerde symposium over 'Research on metacognitive judgements and assessments', waarin ingezoomd werd op de laatste vraag, nl. de wetenschappelijke waarde en de praktische bruikbaarheid van eenvoudige instrumenten voor het meten van 'knowledge monitoring' zowel op macro- als microniveau.

In een symposium over 'The design of learning environments: International perspectives' (georganiseerd als een 'joint session' door Division C en het AERA International Relations Committee) werden door Goldman (namens de Cognition and Technology Group at Vanderbilt), Mandl en Vosniadou (Hatano was

afwezig) leeromgevingen voorgesteld die gebaseerd zijn op recente inzichten over effectief leren en onderwijzen. Tijdens de discussie werd door Collins gewezen op de grote diversiteit in de principes die aan het ontwerp van deze verschillende omgevingen ten grondslag liggen. Derhalve werd ervoor gepleit om in de nabije toekomst de discussie en het onderzoek toe te spitsen op het vergelijken van de diverse systemen en het zoeken naar de écht kritische dimensies van krachtige leeromgevingen.

Evenals de voorbije jaren was ook op dit AERA-congres het aanbod op het gebied van het leren en onderwijzen van wiskundige begrippen, vaardigheden en houdingen zeer ruim vertegenwoordigd. Zelfs wie uitsluitend in dit thema geïnteresseerd was, diende bijna voortdurend een keuze te maken uit twee of zelfs drie parallel geprogrammeerde sessies. Met uitzondering van de spectaculair toegenomen belangstelling voor de problematiek van 'authentic assessment' - waarover meer in de bijdrage van Knoors en Dochy - vallen weinig nieuwe ontwikkelingen te melden in vergelijking met bijv. de stand-van-zaken geschetst in onze bijdrage tot de AERA-kroniek van 1994.

Dit betekent dat de onderzoekers van het reken/wiskunde-onderwijs sedert enkele jaren massaal overgestapt zijn op construerend onderzoek, gericht op het ontwikkelen en uitproberen - in nauw contact met practici - van nieuwe vormen van reken/wiskundeonderwijs. Zo zijn onderzoekers zoals Carpenter, Fuson, Silver, Resnick, Cobb en Lampert allemaal nauw betrokken bij de optimalisering van het nationale reken/wiskunde-onderwijs via het construeren, implementeren en evalueren van nieuwe curricula, deelleergangen, leermaterialen, interventietechnieken, evaluatie-instrumenten, opleidings- en navormingsprogramma's. In al deze studies worden de inmiddels zeer bekende *Standards* (National Council of Teachers of Mathematics, 1989) - een document dat qua strekking en impact vergelijkbaar is met de Nederlandse *Proeve van een nationaal programma voor het reken/wiskunde-onderwijs op de basisschool* (Treffers, De Moor & Feijs, 1989) - uitdrukkelijk en expliciet als referentiekader gebruikt. Bij het overzien van dit onderzoek hebben wij de indruk dat daar noch vanuit theoretisch oogpunt noch

vanuit praktisch oogpunt heel erg veel nieuws te rapen valt voor wie bekend is met de rijke Europese traditie inzake onderzoek en onderwijs van de wiskunde. Een bijkomend probleem is dat al deze Amerikaanse vernieuwingsprogramma's in afzonderlijke sessies voorgesteld en besproken worden, zonder dat ingegaan wordt op eventuele interessante accentverschillen wat betreft achterliggende visie en/of concrete uitwerking.

De socio-culturele context van wiskunde-leren en -onderwijzen blijft een zeer actueel onderzoeksterrein. Ten eerste werden er weerom heel wat papers en posters gepresenteerd waarin ingegaan werd op het informele rekenen van kinderen en volwassenen in buitenschoolse contexten ('ethnomathematics') en op de relatie daarvan met en de betekenis ervan voor het schoolse rekenen. Tijdens deze sessies werd echter nog maar eens duidelijk hoe moeilijk het is om de resultaten van deze op zich boeiende analyses van eigen-aardige manieren van (leren) rekenen te benutten bij de optimalisering van het geïnstitutionaleerd wiskunde-onderwijs.

Internationaal vergelijkend onderzoek is een tweede categorie van onderzoek waarin het socio-culturele aspect van wiskunde-(onderwijs) tot uitdrukking komt. Naast de 'klassieke' comparatieve studies van de wiskundige oplossingsvaardigheden en leer- en denkprocessen van Amerikaanse en Aziatische leerlingen en studenten, werd ook vergelijkend onderzoek gepresenteerd waarin andere landen en culturen betrokken waren. Een goed voorbeeld daarvan was het symposium georganiseerd door Reys over 'Assessing the development of number sense: Collaborative initiatives in four countries', waaraan naast Amerikaanse en Chinese ook Australische en Europese onderzoekers deelnamen. Doch veel verder dan een gezamenlijk referentiekader voor bestudering van 'number sense' en een bruikbare itempool was men nog niet gekomen.

De socio-culturele context van het rekenonderwijs kwam tenslotte ook ruimschoots aan bod in een aantal symposia, papers en posters rond 'the culture of the mathematics classroom'. In dit verband verwijzen we naar ons eigen symposium rond 'Real world knowledge and mathematical modeling of word problems

in schools', met bijdragen van De Corte, Greer, Reusser, Säljö en Verschaffel en een discussiebijdrage van Kilpatrick over de spanning tussen het oplossen van vraagstukken in de schoolse context enerzijds en het oplossen van reële problemen anderzijds.

Tenslotte kwam ook de rol van informatietechnologie (computers, interactieve video) aan bod in heel wat presentaties en demonstraties rond wiskundeleren en -onderwijzen. Bijzonder boeiend was een sessie georganiseerd door Staub, waarin een tiental computerondersteunde 'representational tools for mathematics instruction' (met o.m. bijdragen van Reusser, Derry, Resnick, Greeno, Nathan) werden gepresenteerd en gedemonstreerd. In zijn discussie wist Kaput de verschillende systemen te plaatsen op een aantal kerndimensies (mate van controle, mate van ingebouwde intelligentie, flexibiliteit van de verbanden tussen de diverse representaties). Tevens formuleerde hij een aantal belangrijke uitdagingen waarvoor ontwikkelaars van dergelijke computerleeromgevingen staan. Als belangrijkste wees hij erop dat men cruciale beslissingen dient te nemen over welke aspecten van het wiskundig handelen men kan automatiseren en welke aspecten men de leerling of student best nog zelf laat/leert doen.

Literatuur

- Phillips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Researcher*, 24(7), 5-12.
- National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Treffers, A., Moor, E. de, & Feys, E. (1989). *Proeve van een nationaal programma voor het reken-wiskundeonderwijs op de basisschool. Deel 1. Overzicht einddoelen*. Tilburg: Zwijzen.

C - II: Onderwijstechnologie en ontwerpkunde (J. Elen en J. Lowyck, Katholieke Universiteit Leuven)

Alhoewel onderwijstechnologie en ontwerp-kunde duidelijk aan elkaar zijn gerelateerd, wordt in de verslaggeving eerst ingegaan op de onderwijstechnologie en vervolgens op de ont-

werpkunde. Afsluitend worden enkele trends genoemd die vanuit de, weliswaar beperkte, selectie uit het aanbod onderkend kunnen worden.

Onderzoek binnen de *onderwijstechnologie* lijkt thans sterk georiënteerd op telecommunicatie. Het begrip 'telecommunicatie' is echter een algemene en omvattende term waaronder zowel het raadplegen van 'homepages' als het uitwisselen van boodschappen over 'het net' vallen. Zo beschouwd verwijst telecommunicatie naar een stadium in de technologische evolutie waarin vrijwel alle gangbare onderwijsleerfuncties via genetwerkte computers (zullen) kunnen worden gerealiseerd. Niet alleen netwerken, maar ook de beschikbare software voor multimedia kwam frequent aan de orde. Veel van de gepresenteerde papers benadrukken, zij het op een zeer descriptieve en hypothetische wijze, de invloed van deze ruime informatie-, spel- en leerbronnen op de positie van de school als instituut. Vooral de relatie tussen 'school and community' en 'school and home' zijn voltreffers in het prospecteren van 'the school of the future'. Een voorbeeld hiervan was te vinden in de paper van Mann die wees op de enorme vlucht die de multimedia industrie neemt inzake 'edutainment'. Op de AERA-conferentie was nog duidelijker dan voorheen de kloof merkbaar tussen enerzijds goed gefocuseerd hypothesetoetsend onderzoek en anderzijds ruwe prospecties over de toekomst van het onderwijs en de rol van de onderwijstechnologie daarin. Blijkbaar krijgt de laatste trend op de AERA-conferentie langzamerhand de meeste aandacht.

Laten we de meer speculatieve bijdragen buiten beschouwing, dan lijkt het onderwijs-technologisch onderzoek zich op een dubbel spoor te bevinden.

Enerzijds bestudeert men, in de lijn van vroeger onderzoek, het belang van symbolen-systemen (tekst, beelden, beweging) waarmee de werkelijkheid wordt gerepresenteerd. De klemtoon ligt dan op het concretiseren en manipuleren van informatie met het oog op het verwerven van kennis (Entwistle spreekt, in dit verband over 'knowledge objects'). Parallel hiermee wordt ervoor gepleit om ook in onderwijsleersituaties het (leren) manipuleren van '(kennis)objecten' centraal te stellen. Enyedy

bijvoorbeeld rapporteerde onderzoek naar de effecten van een computeromgeving op het domein van de waarschijnlijkheidsleer. Hij constateerde dat manipuleerbare representaties bijdragen tot de bespreekbaarheid ervan en dat studenten door representaties geholpen worden om actief met die abstracte materie om te gaan. Schnotz op zijn beurt heeft systematisch onderzoek gepresenteerd over bewegende beelden. Hij stelde vast dat het beste leerresultaat wordt bereikt door leerlingen bij wie visueel-spatieel relaties in figuren overeenkomst vertonen met semantische relaties tussen gepresenteerde kenniselementen. Dit helpt verklaren waarom dynamische figuren een beter leereffect sorteren indien ook de te verwerven kennis procesmatig van aard is.

Anderzijds kwam nieuw onderzoek aan bod over implicaties van de nieuwe technologische mogelijkheden voor het leren. Interessant in dit verband is het onderzoek dat zowel aan de University of Illinois als aan het Education Development Center is verricht naar mentale modellen die leraren construeren over 'Internet' en naar de metaforen die ze hierbij hanteren. Toch blijven de bevindingen op dit vlak eerder opsommend en beschrijvend, zonder specifieke aandacht voor het ontstaan of de evolutie van die modellen of voor de invloed ervan op het gebruik van netwerken in onderwijsleersituaties. In de lijn van de trend naar 'netwerken', passeerde een aantal communicatie-omgevingen de revue die leerlingen/studenten in staat moeten stellen met behulp van CMC (computer mediated communication) op afstand met elkaar samen te werken. Zo stelt de communicatie-omgeving KIE (Knowledge Integration Environment), ontwikkeld in Berkeley en gepresenteerd door Bell en Davis, leerlingen in staat verschillende soorten van evidentie op te zoeken, te rapporteren en vervolgens te bespreken. Doordat leerlingen in KIE ook expliciet worden aangespoord hun opmerkingen en theoretische uitgangspunten te bespreken, wordt verwacht dat ze op een dieper niveau met de materie bezig zijn en hierdoor een dieper inzicht verwerven. Kaye ging nader in op de kenmerken van effectieve netwerk-omgevingen en meer bepaald op 'Model Nets' als systeem voor computer netwerking vanaf de kleuterschool tot het einde van het secundair onderwijs (K-12). Helaas werd geen serieuze evaluatie

van de leereffecten gerapporteerd, maar wel de optimistische conclusie dat schooldistricten die een dicht netwerk creëren, ook bijdragen tot de verbetering van de kwaliteit van hun onderwijs. Het gehanteerde criterium voor deze voorbarige conclusie was uitsluitend de mate van penetratie van computer netwerken. Dat het gebruik van telecommunicatie wel de organisatie van het onderwijs beïnvloedt maar nauwelijks effecten te zien geeft op individuele leerresultaten werd onder andere geïllustreerd in een paper van Hawkins en Honey 'Social factors for networking'. Op grond van ervaringsgegevens werd gewezen op het belang van passende ondersteuning. Omdat deze ondersteuning door computers gemedieerd wordt, dient volgens de onderzoekers deze typische wijze van communiceren vanaf de start in rekening te worden gebracht. Deze gemedieerde ondersteuning riep verder nog vragen op over de behoefte van persoonlijke aanwezigheid van de ondersteunende instantie en over de mate waarin gemedieerde ondersteuning als authentiek en betrouwbaar wordt ervaren door lerenden. Een hamvraag in het startend onderzoek naar effecten van telecommunicatie, is de evaluatie. Zowel Stecher als Quellmalz en Kozma stelden de vraag naar de wijze waarop deze nieuwe mogelijkheden kunnen worden geëvalueerd. Laatstgenoemden presenteerden een complex schema dat echter niet wezenlijk afwijkt van een gangbaar evaluatieschema. Dit wijst op de moeilijkheid om nieuwe omgevingen grondig te analyseren in hun typische maar ladende kenmerken. Volgens Stecher wordt degelijke evaluatie ook bemoeilijkt door de complexiteit die ontstaat door een 'mix' van activiteiten en technologieën, die dankzij de telematica parallel kunnen verlopen op verschillende locaties die wel aan elkaar gerelateerd zijn maar niettemin onafhankelijk van elkaar kunnen opereren.

Een volgende stap in de implementatie van nieuwe technologieën is het aanwenden van 'virtual reality' (VR), waarover een symposium werd georganiseerd. Het systematisch onderzoek naar de mogelijkheden en grenzen ervan voor het onderwijs staat duidelijk nog in de startblokken. Wel zijn tal van evaluatieonderzoeken gaande met soms positieve uitkomsten. Dede ging in op het gebruik van VR voor het bevorderen van onderwijs in de na-

tuurwetenschappen. Hiertoe werden diverse virtuele werelden gecreëerd waarin lerenden zich zelfstandig in een driedimensionele virtuele simulatie-omgeving kunnen bewegen om verschillende hypothesen uit te testen. Het gebruik van VR wordt door Dede beargumenteerd vanuit de mogelijkheden tot driedimensionele representatie, het gebruik van meerdere sensorische systemen, en tot 'immersion' en motivatie, eigen aan VR. Terecht stelde hij echter dat onderwijskundig nut hier voorrang verdient, wat niet steeds evident is. VR wordt dan niet zozeer als een omvattend antwoord, maar als een hoogtechnologische en specifieke mogelijkheid naar voren gebracht. De kosten voor het ontwikkelen van VR-omgevingen bracht wellicht mee dat in dit symposium verder alleen de NASA en de Landmacht van de VS aan het woord kwamen. Loftin (NASA) ging nader in op een project waar astronauten werden voorbereid op hun missie voor het herstellen van de Hubble-telescoop. Daarbinnen werd ook tussen astronauten samengewerkt die zich op ver van elkaar gelegen plaatsen bevonden. Goldberg beklemtoonde dat het gebruik van VR voor het leger kostenbesparend kan zijn door de behoefte aan training te reduceren en te verhinderen dat reële uitrusting moet worden gebruikt. Voor hen vormt VR een uitbreiding van gangbare simulaties. In de laatste twee papers werd gewezen op zogenaamde 'simulator sickness' en werd de vraag gesteld naar de mate waarin VR de werkelijkheid getrouw moet afspiegelen om voor de lerende betrouwbaar over te komen.

Met betrekking tot *ontwerpkunde* vallen onmiddellijk twee zaken op. Ten eerste is 'instructional design' als apart onderzoeks- of discussieterrein vrijwel verdwenen. De vroegere debatten over de relatie constructivisme-objectivisme zijn nagenoeg uitgedoofd. Voor zover we konden nagaan was er slechts één groep van paperpresentaties waar in de titel expliciet naar ontwerpkunde werd verwezen en waar overigens weinig nieuws te rapen viel. Zo rapporteerde Maxwell over eigen ervaringen naar aanleiding van zijn poging om binnen een MBA-programma ontdekkingsleren te implementeren. Wedman rapporteerde een studie die het belang van 'analogical problem solving' reveleerde via hardop denken. Vooral de gehan-

teerde methode voor het analyseren van de hardop-denken protocollen lijkt veelbelovend. Er moet worden geconcludeerd dat het onderzoek naar ontwerpen als proces stilgevallen is en het onderzoek naar meer formele ontwerpproducten ondergeschikt is gemaakt aan domeinspecifieke of media-gerelateerde vraagstellingen. Toch dienen nog twee symposia te worden vermeld die dit beeld enigszins relativeren. Vanuit de 'SIG Structural Learning' werd een Amerikaans-Europees symposium georganiseerd over het automatiseren van het ontwerpen. Dit is zowat een traditie geworden op de AERA conferentie, zoals uit de programma's van vele jaren blijkt. Er werd voornamelijk aan taakanalyse aandacht besteed. Zo presenteerde Scandura een gecomputeriseerd hulpmiddel om taken te analyseren en rekening houdende met de uitkomsten ervan opleidingen te ontwerpen. Tennyson demonstreerde een recente versie van GAIDA, GUIDE genaamd, en bepleitte het construeren van een set van losse design tools, in tegenstelling tot een meer geïntegreerd systeem zoals de 'ID-Expert' van Merrill c.s., dat het ontwerpen van courseware volledig zou automatiseren. Op de vraag naar de behoefte aan onderlinge samenhang van een set losse ontwerptools werd door Tennyson niet ingegaan. Dijkstra stelde een systeem voor om doelen en leertaken te categoriseren, als een degelijke basis voor het nader analyseren van leertaken. Schott beklemtoonde eveneens het belang van taakanalyse en lichtte een in Dresden ontwikkelde cognitieve taakanalyse-procedure toe.

Directer met de ontwerpkunde verbonden was het symposium van de University of California in Berkeley. Binnen de onderzoeksgroep waaraan ook Lynn en diSessa participeren wordt expliciet gestreefd naar het formuleren van een prescriptieve instructietheorie. Gesteld werd dat de theorie van Bruner vervangen dient te worden door een theorie die de technologische mogelijkheden van computers mee in rekening brengt. Alle inspanningen binnen de onderzoeksgroep zijn georiënteerd op het formuleren van zo'n instructietheorie. Eerder dan losse fragmenten vormen de verschillende onderzoeken binnen de onderzoeksgroep (bijv. naar communicatie-omgevingen (Bell & Davis), samenwerken (Hoadley & Hsi) en simulatie-omgevingen (Enyedy)) bouwstenen

voor de nog te construeren onderwijstheorie waarin de technologie volledig tot zijn recht komt. Centraal staat hierbij, volgens Gifford, de gedachte dat technologie een integraal bestanddeel kan zijn van een onderwijsleeromgeving en dat leren, ondersteunen van leerprocessen en het ter beschikking stellen van informatie, vooral functies zijn en niet zozeer aan personen of objecten gebonden opdrachten. Wat het kennisdomein betreft, ging veel aandacht gaan naar de mogelijkheden van technologie voor het ondersteunen van de verwerving van wetenschappelijke inzichten.

In een aantal onderzoeken ging aandacht uit naar een vraagstelling die steeds meer met ontwerp-kunde in verband wordt gebracht, namelijk de percepties van studenten en de eventuele invloed ervan op het leren. Hiermee verbonden is de kloof tussen percepties van studenten en onderwijsgevend en ontwerpers. Zo benadrukte Reiman in zijn analyse van onderzoek naar voorkennis, de rol van 'task perception'. Reusser presenteerde een poster over de invloed van percepties over groepswork en Lowyck rapporteerde tijdens het symposium over het automatiseren van ontwerpen een Leuvense studie (Elen & Lowyck) naar de kennis die studenten bezitten over diverse - ook technologische - leeromgevingen. Inzicht in deze mediërende variabelen kan bijdragen tot effectievere leeromgevingen door met het 'gebruikersstandpunt' meer expliciet rekening te houden tijdens het ontwerpen en uitvoeren van instructie.

In t al van symposia kwamen thema's aan de orde die voor het ontwerpen van onderwijsleeromgevingen belangrijke vragen oproepen. In dit verband kan eerst en vooral verwezen worden naar het symposium over 'Conceptions of learning and understanding' waarover Verschaffel en De Corte reeds gerapporteerd hebben.

Door tal van onderzoekers wordt gezocht naar een nieuwe conceptualisering van factoren die het leren beïnvloeden. Vanuit een sociaal constructivistisch standpunt werd door McCarthy het argument aangebracht dat de determinerende factoren, resp. de voor leren interessantste variabelen zich niet bevinden in de (kenmerken van) interagerende actoren maar in de (aard van de) interactie zelf. In dit perspectief

wordt de vraag naar de aard van de interacties die een leeromgeving toelaat belangrijker dan vragen naar de mate waarin een individueel lerende door deze omgeving wordt ondersteund.

Iets breder terugblikkend en op grond van symposia, papers en posters die ook buiten de strikte ontwerp-kunde en onderwijstechnologie vallen, kunnen de volgende tendensen worden onderkend die voor de toekomst van (het onderzoek binnen de) onderwijstechnologie en ontwerp-kunde van belang kunnen zijn.

- * Er is verhoudingsgewijze weinig systematisch onderzoek voorhanden dat beoogt modellen en (deel-)theorieën voor het domein van de onderwijstechnologie en ontwerp-kunde te ontwikkelen.
- * Veel onderzoek in het kader van de onderwijstechnologie richt zich op organisatorische kwesties, en veel minder op leerprocessen en -uitkomsten.
- * Op het micro-niveau van het leren, wordt in toenemende mate een rol toegekend aan biologie en neurologie. Farley baseert zijn T-type - typologie relevant in verband met creativiteit - op de bevinding dat één gen verantwoordelijk zou zijn voor interindividuele verschillen qua 'risk-taking'. In dezelfde lijn beklemtoonde Pintrich in zijn discussie over de stand van zaken in het onderzoek naar voorkennis de behoefte aan 'biological metaphors'.
- * Er blijkt een steeds terugkerende vraag te zijn naar stabiele raamwerken en ankerpunten. Deze zoektocht leidt tot een aanpak waarin de gehele context in zijn complexiteit wordt betrokken, dan wel tot veel aandacht voor persoonsgebonden variabelen ('belief systems', 'task perceptions', 'personality') en hun invloed op het leren. Deze dubbele tendens leidt enerzijds tot een niet-geproblematiseerd gebruik van concepten die de omgeving beschrijven, zoals 'context' en 'culture', en anderzijds tot het beklemtonen van de rol van variabelen als 'personality', 'gender' of erfelijkheid.
- * Er wordt nauwelijks replicatie-onderzoek verricht. In plaats van zich in de lijn van een traditie te plaatsen, gaat de voorkeur uit naar onderzoek dat als nieuw, origineel of als een uiting van de zeer recente paradigma-wisseling wordt bestempeld. Door het af-

wijzen van continuïteit in het onderzoek wordt het erg moeilijk om onderzoeksgegevens te accumuleren tot op het niveau van model- en theorievorming. Het gebrek aan stabiliteit en continuïteit lijkt daarom voor de uitbouw van een solide ontwerp-kunde en onderwijstechnologie problematisch.

- * Oorspronkelijk zeer technische psychologische en/of sociologische begrippen of omschrijvingen van stromingen worden geleidelijk geabsorbeerd en op een algemeen niveau getild waardoor ze al te omvattend en betekenisloos worden. Een dergelijke evolutie vond reeds vroeger plaats, bijvoorbeeld voor de concepten 'self-regulation' en 'constructivism' waarop alle mogelijke betekenissen werden geprojecteerd. Thans lijkt dit het geval voor termen als 'representation', 'mental model', en zelfs 'distributed cognition'.

Bij wijze van conclusie kan gesteld worden dat de AERA-conferentie opnieuw gewezen heeft op de spanning tussen het zoeken naar stabiliteit en algemeen geldende regels enerzijds en het erkennen van de complexiteit en eigenheid van ieder onderzoeksobject anderzijds. Niet voor niets spreekt bijv. Pintrich in een symposium over zelfregulatie over 'a cludge' en 'the rich messiness of learning'.

Divisie D: Measurement and Research Methodology (E. Knoors, Katholieke Universiteit Leuven en F. Dochy, Open universiteit Heerlen)

Het zal duidelijk zijn dat Divisie D te veel omvat om alles hier te beschrijven. Wel kunnen we enkele algemene trends aangeven. Daarna geven we een overzicht van enkele door ons bijgewoonde presentaties en workshops.

Algemene trends

In totaal werden er in deze divisie ongeveer 75 sessies georganiseerd. Hierin is de laatste jaren een steeds groter aandeel te zien van onderwerpen in verband met 'assessment' en 'measurement', gepaard gaande met een krimp van de zuiver methodologische en psychometrische sessies.

Veel voorkomende onderwerpen zijn 'Performance Assessment', 'Measurement of

Change', 'Assessment-Based Educational Reform', 'Embedding Assessment systems', 'Collaborative assessment', 'Assessment of Growth', 'Growth modeling', 'New forms of Assessment', etc. Ook het slechts één keer opduiken van de term 'betrouwbaarheid' in de programma-index, tegenover de vele sessies over 'validiteit', vormt een aanduiding voor deze ontwikkelingen. We lichten deze trend kort toe.

De laatste jaren is er sprake van een verschuiving van een test-cultuur naar een 'assessment'-cultuur; van een (puur) psychometrische aanpak naar een aanpak waarin de aandacht vooral uitgaat naar de authenticiteit van het testmateriaal; van een aanpak waarin primair de aandacht uitgaat naar betrouwbaarheid, naar een aanpak waarbij validiteit en 'fairness' centraal staan. Door de al te grote aandacht voor de 'high-stakes testing programs' (dit wil zeggen, prestatietoetsen met belangrijke consequenties voor de leerling) is men tot het inzicht gekomen dat toetsing in eerste instantie voor de student bedoeld is. Hij moet het nut ervan inzien en er iets mee kunnen. Daarnaast heeft onderzoek de grote invloed aangetoond van toetsing op het leren van studenten en hun studiegedrag in het algemeen. Daarmee gepaard ging de stijgende bewustwording van het niet-authentieke karakter van examens en het verlies aan geloof hierin als valide toetsen van leerresultaten. Klassieke testen blijken vaak niet eerlijk te zijn voor de student. De reactie hiertegen wordt steeds sterker. Het is belangrijker valide, authentieke toetsen te ontwikkelen die eerlijk zijn voor studenten en die via meerdere methoden worden afggenomen. Men beklemtoont steeds meer het belang van formatieve toetsen, het bewaken van de voortgang van het leerproces, en van het geven van duidelijke, informatieve feedback.

Workshops

Naast presentaties werd er ook een aantal minicursussen en 'pre-conference courses' over onderzoeksmethoden georganiseerd. Op basis van hun jarenlange ervaring met PLSPath verzorgden Keeves en Campbell een 'preconference session' van één dag getiteld "Soft Modeling - Path Analysis 'Lite'". Hierin werd een software programma voor pad-analyses gepresenteerd - PLSPath - dat zich van andere on-

derscheidt doordat het gebaseerd is op een 'least squares' aanpak en vrij is van onderliggende assumpties. Het programma werd voorgesteld als krachtig maar eenvoudig genoeg om met een minimale training en voorkennis te gebruiken. Ten minste, als je weet hoe je de resultaten moet interpreteren. Een minpunt van het programma is de grote hoeveelheid output en het feit dat het om een DOS-programma gaat. Duidelijk was dat ook de instructeurs geen gebruik maakten van alle detailgegevens die de analyses opleveren, maar zich enkel beperken tot de 'jackknife' procedures.

Een drie uur durende cursus over 'Interrater Reliability in Performance-based Assessment Measures' werd verzorgd door Abedi. Over het reduceren van variantie, veroorzaakt door de scoorders, werd weinig nieuws verteld: Men dient concrete en valide scoringsrubrieken te voorzien en de scoorders moeten goed getraind zijn. Voor het schatten van de interbeoordelaarsbetrouwbaarheid werden acht relatief eenvoudige methoden voorgesteld, variërend van een Product-Moment Correlatie tot een 'William's index of agreement'. De vergelijking van deze diverse waarden voor 'interrater reliability' geeft een grotere mate van zekerheid omtrent de consistentie tussen scoorders, ook al vertrekken deze methoden van een verschillende basis. Abedi legde ook uit dat het gebruik van meerdere methoden om 'interrater reliability' te schatten van belang is, omdat ze beïnvloed worden door verschillende aspecten van de studie, zoals het aantal scoorders en de grootte van de steekproef.

Opvallend, en in contrast met de relatief eenvoudige maten uit Abedi's sessie, is dat in een sessie over 'Performance assessment scoring issues' diverse meer complexe methoden en analyses werden gepresenteerd om verschillen tussen beoordelaars aan het licht te brengen. Twee presentaties in deze sessie verdienden zeker de aandacht. Zowel Martin als Clouser gaven een goede presentatie over het geautomatiseerd scoren van complexe 'performances'. Het is een goede ontwikkeling dat er verder nagedacht wordt over het automatiseren van het scoren van 'performance assessment'. Voor toetsontwikkelaars, en zeker ook voor de opdrachtgevers daartoe, is de hoeveelheid tijd en moeite die gepaard gaat met het scoren van 'performance assessment' veelal een belem-

mering, vooral als men dergelijke toetsvormen op grote schaal wil toepassen. Martin en Clouser maakten dan ook duidelijk dat het automatiseren van het scoren van 'performance assessment' alleen zin heeft als de studentenaantallen groot zijn. In hun geval ging het over een artsenexamen dat voor tienduizenden subjecten moest dienen.

'Performance assessment – Alternative assessment'

Als we de titels van alle sessies in deze divisie en de onderwerpindex van het programmaboek bekijken, zien we dat 'performance assessment' op AERA '96 erg in de belangstelling stond. Bij een tiental sessies was 'performance assessment' in de titel opgenomen, terwijl niet minder dan 38 sessies in de index onder 'performance assessment' gerangschikt stonden. In de presentaties over dit onderwerp kunnen eigenlijk twee hoofdthema's onderscheiden worden. Een eerste thema betreft de relatie tussen toetsing en instructie: In welke mate kunnen toetsen richting geven aan het onderwijs in de klas, hoe kan toetsing geïntegreerd worden met het curriculum, en welke ervaringen zijn er met betrekking tot het invoeren van nieuwe toetsvormen? Een tweede thema betreft de vraag hoe aan dergelijke instrumenten voor 'performance assessment' vorm gegeven moet worden: Welke toetsvorm is geschikt, hoe moeten de items eruit zien, hoe zit het met de betrouwbaarheid en validiteit van 'performance assessment'?

Alvorens enkele interessante presentaties in verband met deze twee thema's te beschrijven, wijzen we er nog op dat het onderzoeks- en ontwikkelingswerk in verband met 'performance assessment' heel sterk toegespitst is op het gebied van wiskunde en taal. In het vervolg van dit stuk zullen we exemplarisch ingaan op 'performance assessment' binnen het gebied wiskunde, omdat de meeste door ons bijgewoonde presentaties daarop betrekking hadden.

De enorme belangstelling voor de kwestie van 'assessment' bij onderzoekers en ontwikkelaars van het wiskunde-onderwijs (zie de bijdrage van Verschaffel & De Corte) houdt verband met het groeiend besef dat elke vernieuwing van dit onderwijs tot mislukking is gedoemd, indien die niet gepaard gaat met of zelfs geleid wordt door parallelle ontwikkelin-

gen inzake toetsontwikkeling en -gebruik. In verscheidene sessies had men het in dit verband over 'assessment-based educational reform'.

Een van de opvallendste karakteristieken van 'performance assessment' is dat items vaak ingebed zijn in authentieke of realistische situaties. In een sessie getiteld 'Investigating relationships between student learning and assessment in primary schools', stelde Cooper zich echter de vraag of dit de eerlijkheid van de toetsing ten goede komt. Op basis van een pilootstudie met een aantal 10- en 11-jarige leerlingen beargumenteerde hij dat de sociale klasse waartoe kinderen behoren, verschillen teweeg kan brengen in de manier waarop zij toetsitems interpreteren en begrijpen. Op basis van de resultaten van interviews die hij met de leerlingen hield enerzijds en gebruikmakend van theorieën van Bernstein en Bourdieu anderzijds, kwam hij tot de (voorlopige) conclusie dat kinderen uit de arbeidersklasse hun oplossingsstrategieën en antwoorden meer baseren op contextgebonden ervaringskennis dan kinderen uit de middenklasse, die gebruik maken van abstracte, wiskundige kennis. Het interessante aan Coopers studie is dat daaruit blijkt dat het inschakelen van deze concrete ervaringskennis deze leerlingen hindert bij het oplossen van wiskundige toepassingsproblemen, omdat zij die aanwenden op plaatsen waar eigenlijk wiskundige kennis vereist is. In die zin vormt Coopers analyse een interessante aanvulling op de studies die in het symposium over 'Real-world knowledge and mathematical modelling of word problems in schools' (zie de bijdrage van Verschaffel & De Corte) werden besproken. Cooper droeg aan het slot van zijn betoog dan ook de waarschuwing uit dat wanneer deze conclusies in verder onderzoek bevestigd worden, toetsontwikkelaars voor moeilijke beslissingen komen te staan omtrent het gebruik van toetsopgaven die aan de ene kant het onderwijs kunnen verbeteren, maar aan de andere kant leiden tot minder eerlijke toetsresultaten.

In dezelfde sessie wees ook Harlen op mogelijke negatieve invloeden van 'assessment' op het leren. Om juist een positieve invloed van 'assessment' op het leren te bewerkstelligen, is het volgens Harlen belangrijk dat het formatief toetsen zo gedefinieerd wordt, dat het duidelijk te onderscheiden is van summatief toetsen. Pas

als de essentiële verschillen blootgelegd zijn, kunnen voorstellen gedaan worden om beide toetsvormen aan elkaar te koppelen. De assumpties (a) dat formatieve toetsen *niet* wezenlijk verschillen van summatieve toetsen, en (b) dat de resultaten van een formatieve toets ook *zonder meer* geschikt zijn voor een summatieve evaluatie, zijn namelijk bronnen voor een negatieve impact van toetsen op het leren. Men moet zich realiseren dat de informatie die leerkrachten opdoen tijdens hun onderwijsactiviteiten fragmentarisch en vaak tegenstrijdig is. Deze incongruente gegevens zijn voor een formatief doel goed om aan te duiden wat een leerling vooruit helpt of juist remt in zijn leerproces. Harlen benadrukt echter dat voor een summatieve evaluatie weliswaar hetzelfde materiaal gebruikt kan worden, maar dat de gegevens in het licht gesteld moeten worden van de criteria die voor alle studenten gelden. Dit betekent dat het complete werk van een leerling holistisch bezien moet worden, waarna beoordeeld wordt of het daaruit voortvloeiende profiel voldoet aan de criteria.

Met betrekking tot het tweede thema, namelijk de vormgeving van toetsinstrumenten, werd veelvuldig de vraag gesteld welk type opgaven geschikt is om leerprocessen en vaardigheden van hogere orde in beeld te brengen. In het algemeen was er een trend waar te nemen tegen gestandaardiseerde toetsen en, meer algemeen, tegen geschreven, korte-antwoord items. In het symposium 'Developing valid large-scale mathematics performance assessments that provide diagnostic information' beschreef Close - net als Cooper in de eerder genoemde sessie - hoe in Engeland de toetsontwikkelaars zich afkeerden van dergelijke toetsen en de weg insloegen naar meer 'authentieke' toetsen met complexe testitems; doch onder druk van de regering dienden ze terug te keren naar de klassieke 'paper-and-pencil' toetsen. Van den Heuvel-Panhuizen verzette zich dan weer tegen auteurs als Joffe en Clements, die toetsen met geschreven, korte-antwoord opgaven per se overboord willen gooien. Zij beargumenteerde dat dergelijke toetsen niet uitgebannen kunnen en moeten worden. Men zou volgens haar beter energie kunnen steken in het verbeteren van deze toetsvorm, zodat de negatieve kenmerken en invloeden ervan geëlimineerd worden. Daartoe moe-

ten wel een aantal taboes doorbroken worden die lange tijd de vormgeving van dergelijke toetsen bepaald hebben, zodat men bijvoorbeeld aanvaardt dat een wiskundig probleem meer dan één goed antwoord kan hebben en dat het antwoord op een probleem soms niet exact te bepalen is. (In dit verband merkte Lesh in zijn discussiebijdrage op dat "there is no right way to be good in math"). Dan blijkt dat ook geschreven, korte-antwoord opgaven en zelfs meerkeuzevragen zeer informatief kunnen zijn over de leerprocessen, strategieën en hogere denkvaardigheden van leerlingen.

Ook in de sessie getiteld 'Issues in math assessment' stond het tweede thema, namelijk de vormgeving van 'performance assessment', centraal. Zoals eerder opgemerkt, is de tijd die gemoeid is met 'performance assessment' veelal een probleem. Misschien is het daarom dat Suzuki zich de vraag stelde wat het minimum aantal items is in 'performance based tests'. Met behulp van diverse statistische technieken is ze de betrouwbaarheid van toetsen met verschillende lengten nagegaan, en op basis daarvan concludeerde zij dat reeds met negen opgaven een betrouwbare toets samengesteld kan worden; mits ze zorgvuldig gekozen worden kunnen vijf items zelfs al voldoende zijn. Hoewel het op zich goed nieuws is dat ook een toets met zo weinig opgaven al een betrouwbare meting kan opleveren, slaagde Suzuki er onvoldoende in te expliciteren wat de aard van dergelijke opgaven moet zijn. Bovendien merkte Ankenmann als discussant op dat een toets met zo'n klein aantal items altijd in te boeten heeft aan validiteit. Eerder is reeds aangeduid dat de validiteit van toetsen in een stijgende belangstelling staat, tegenover een overwaardering van de betrouwbaarheid in het verleden. Deze eenzijdige aandacht voor de betrouwbaarheid van een toetsinstrument, waarbij de inhoudsvaliditeit verwaarloosd wordt, is ook Freudenthal een doorn in het oog geweest. Door Brown werd in dezelfde sessie een studie gepresenteerd die zowel de betrouwbaarheid als de validiteit van het door hem gebruikte instrument voor 'performance assessment' van wiskunde, in ogenschouw neemt. Een nog maar eens bevestigde, maar nog te weinig toegepaste bevinding van deze studie is, dat analyses gebaseerd op de item-responstheorie en de klassieke testtheorie heel goed gecombi-

neerd kunnen worden in de toetsconstructie. Tevens kan deze combinatie van methoden aangewend worden om het toetsstelsel in het oog te houden en te verbeteren, onder andere met betrekking tot de betrouwbaarheid en de validiteit. Ons inziens wordt het vaak voorgesteld alsof de item-responstheorie en de klassieke testtheorie in een soort machtsstrijd verwickeld zijn. Wellicht is het blijvend probleem dat een praktijktoepassing van de IRT goed gecalibreerde items vereist. Theoretici propageren de ene of de andere psychometrische methode, terwijl in de studie van Brown de twee een symbiose aangaan. FACETS (het gebruikte IRT model) schaalt alle facetten van 'performance assessment' (studenten, scoorders, items, enzovoort), waardoor het mogelijk is om elementen binnen één facet te vergelijken, alsook verschillende facetten onderling te vergelijken. Door de combinatie van FACETS en de klassieke testtheorie kan de overeenstemming tussen scoorders, de interbeoordelaarsbetrouwbaarheid en de toetsvaliditeit worden onderzocht.

Divisie G: Social Context of Education

(G. W. Meijnen, Universiteit van Amsterdam)

Traditioneel vormen onderwerpen als 'ouderschool', de relatie tussen schoolloopbanen en sociaal-economische respectievelijk etnische herkomst en de relatie met sekse het hart van de divisie. Afgeleide onderwerpen als effecten van tweetalig onderwijs, verwachtingen van leerkrachten, leerlingen en ouders, processen van schoolkeuze en sociale relaties binnen klassen behoren uiteraard ook tot de veelvuldig opgevoerde onderwerpen.

Opvallend is voorts dat Amerikaanse onderzoekers in toenemende mate belangstelling aan de dag gaan leggen voor wat zich in het onderwijs in Japan, China en andere snel expanderende economieën afspeelt. Daarbij gaat het niet alleen om het niveau van de leeropbrengsten maar ook om bijvoorbeeld seksespecifieke curriculuminhouden en om de relaties ouderschool. Uit de aandacht voor de laatste onderwerpen blijkt dat men ook op zoek is naar de inbedding van de instructiecontext in deze landen. Daardoor zal waarschijnlijk duidelijk worden wat wel en niet te gebruiken is voor de Amerikaanse context.

Traditioneel wordt binnen deze divisie ook tijd ingeruimd voor papers die meer beschouwend en al of niet maatschappijkritisch van aard zijn. De kwaliteit van deze lezingen was ook dit jaar zeer wisselend, variërend van erudiete analyses tot pamfletachtige teksten. Kennelijk lukt het de organisatoren niet om zodanige selectieprincipes te hanteren dat een minimumniveau gegarandeerd is.

Hoogtepunten dit jaar voor divisie G waren een optreden van Bernstein en een bijeenkomst gewijd aan de onlangs overleden Coleman. In een tot de nok toe gevulde zaal presenteerde Bernstein op de hem eigen wijze, te weten het prototype van de hoogleraar uit 1950, een nieuwe theorie omtrent elites, onderwijsconcepties en organisatieprincipes. Het leverde een groot aantal omschrijvingen van categorieën op met een zekere herkenningswaarde, maar daar bleef het dan ook bij. De wetenschappelijke verantwoording van de taxonomie was nogal aan de magere kant. Wellicht zal hij de toekomst besteden aan de uitwerking ervan. De op het podium genodigde referenten waaronder Apple en Wexler bereden of hun eigen stokpaardjes of bleven erg vaag in hun respons.

De bijeenkomst gewijd aan Coleman was eveneens zeer druk bezocht. Een aantal prominente collega's waaronder Bidwell en Sörensen refereerden de betekenis van Coleman voor zowel het onderwijssociologisch onderzoek als voor het beleid. Beide zijn onomstreden. Toch ontbraken enige kritische noten niet: zo presenteerde Sörensen een heranalyse van door Coleman gebruikte data om de relevantie van sociaal kapitaal aan te tonen en kwam daarbij tot geheel andere conclusies.

Hoewel qua organisatie niet maar qua onderwerp wel tot Divisie G behorend, was het symposium naar aanleiding van het boek van Herrnstein c.s. *The Bell Curve*, de moeite waard. De drie referenten waren allen van mening dat de inhoud van het hoofdstuk over de beleidsaanbevelingen niet logisch uit de analyses voortvloeit. Een der referenten noemde de wetenschappelijke analyse dan ook het sausje dat de ideologische boodschap verteerbaar moest maken. Wellicht lag het aan de keuze van de panelleden dat het verwachte verbale vuurwerk uit de zaal achterwege bleef.

Samenvattend: het blijft binnen deze divisie een moeilijke opgave bijeenkomsten te selecteren die de moeite waard zijn als het niet om zeer gevestigde namen gaat. En ook dan kan men teleurgesteld worden. Een oplossing ligt niet voor het oprapen.

Divisie H: School Evaluation and Program Development (B. P. M. Creemers, Rijksuniversiteit Groningen)

De eindredacteuren van de kroniek van de AERA-conferentie 1996 verzochten me een bijdrage te leveren over de sessies in Divisie H. Divisie H van de AERA moet het bewijs vormen dat wetenschap en onderzoek ook dienstbaar kunnen zijn voor praktijk en beleid. Het is dan ook een divisie waarin praktijkgericht onderzoek – dat betekent de presentaties van onderzoekers die in de onderwijspraktijk werkzaam of van practici die ook wat onderzoek doen – aan de orde komt. Een dergelijke divisie is bij het plannen van mijn programma voor de AERA-conferentie niet mijn primaire vertrekpunt, maar omdat ook onderwijs-effectiviteit en de verbetering van onderwijs gebaseerd op de resultaten van het onderzoek waarnaar mijn belangstelling uitgaat in deze divisie hun plaats krijgen, was het doornemen van het programma op onderdelen van Divisie H zinvol.

Wanneer men het programma van Divisie H en dat van de 'Special Interest Group School Effectiveness and School Improvement' in beschouwing neemt, dan lijkt het erop dat onderwijs-effectiviteit weer terug is op de agenda van de AERA. Er waren meer presentaties over deze onderwerpen dan in voorafgaande jaren, toen het erop leek dat onderwijs-effectiviteit na een hausse in de belangstelling in de jaren tachtig, in de jaren negentig nauwelijks meer enige interesse kon opwekken. Maar opname onder het kopje 'School Effectiveness and School Improvement' garandeert nog niets over de inhoud noch over de kwaliteit van het gebodene. Bij veel presentaties en in de programmering, zowel bij de SIG 'School Effectiveness and School Improvement' als bij Divisie H, ligt het accent op de onderwijsverbetering, hoewel bij de SIG nog een relatie met effectiviteitsonderzoek aanwezig is.

Over de relatie tussen onderwijs-effectiviteitsonderzoek en onderwijsverbetering is de

afgelopen jaren door diverse auteurs gepubliceerd. Dit gebeurde omdat er, na de aanvankelijk innige relatie, tot uitdrukking komend in deze SIG en in de International Congress for School Effectiveness and Improvement, toch enige distantie tussen beide groepen is ontstaan. Teleurstelling was er aan de kant van degenen die betrokken waren bij onderwijsverbetering vanwege het feit dat er geen pasklare recepten en antwoorden voor problemen in de onderwijspraktijk te geven waren en ook omdat het gewoon af en toe niet lukte; bij degenen die betrokken waren bij onderzoek op dit terrein was er een grotere interesse, naar mijn mening terecht, in de verdere theoretische fundering en beproeving van een integraal model voor onderwijseffectiviteit dan in het opleveren van deze pasklare antwoorden voor de problemen in de onderwijspraktijk, die bovendien vaak ook niet gebruikt werden.

Op basis van de analyse van verschillende perspectieven van effectiviteitsonderzoek en onderwijsverbetering zijn door diverse auteurs voorstellen ontwikkeld voor verbetering van het gebruik van elkaars activiteiten (Creemers & Reezigt, in druk; Reynolds, Hopkins & Stoll, 1993).

In de presentaties op de AERA werd de nadruk gelegd op schoolverbetering, waarbij veelvuldig klachten werden geuit over het onderwijseffectiviteitsonderzoek. Vooral het feit dat het onderzoek te ver van de praktijk afstaat en deeloplossingen aanbood die niet geïntegreerd konden worden in de onderwijspraktijk, de benadrukking van de kwantitatieve methoden voor onderzoek en het belang van de opbrengsten van onderwijs in termen van leerprestaties, vormden daarbij punten van kritiek. Onderwijsverbetering lijkt vooral hand in hand te gaan met andere dan cognitieve opbrengsten en een accentuering van de processen in het onderwijs. Het succes van de onderwijsverbetering wordt dan afgemeten aan de mate waarin veranderingen zijn opgetreden in de factoren en karakteristieken van onderwijseffectiviteit, zoals schoolleiderschap, teamvorming en dergelijke, in plaats van aan de leerwinst van leerlingen.

De presentaties over de onderwijsverbetering betroffen vaak aspecten en factoren die eventueel in de verbetering een rol zouden kunnen spelen, zoals het verlangen van leer-

krachten om te participeren in besluitvormingsprocessen gedurende het veranderings- ('restructuring')-proces (Taylor, Tashakkori & Hardwick, 1996). Bij deze studie waren slechts zeven scholen betrokken, waarbij niet een relatie tussen de wens tot participatie en de participatie zelf kon worden geconstateerd, maar 'gelukkig wel' een duidelijke relatie tussen participatie en positieve opvattingen over participatie. Ook in andere onderzoeken van onderwijsverbetering valt het geringe aantal eenheden op. Er werden verstrekkende conclusies getrokken over zaken als school-based management en onderwijs in grotere tijdseenheden op grond van een gering aantal waarnemingen. Waarschijnlijk wordt dit voor een deel veroorzaakt door de praktijkgerichtheid - er moet een praktijkadvies verstrekt worden - en de betrekkelijk beperkte werksituatie van de onderzoeker.

In de sessies waarin gerapporteerd werd over onderwijseffectiviteitsonderzoek viel in de eerste plaats op dat veel bijdragen van buiten de Verenigde Staten kwamen. Het Nederlandse onderzoek op dit terrein was goed vertegenwoordigd, evenals het Britse onderzoek. Dit zou erop kunnen wijzen dat onderwijseffectiviteit als researchprogramma vooral in Europa tot ontwikkeling gekomen is, terwijl in de Verenigde Staten de nadruk is komen te liggen op onderwijsverbetering, behalve in een aantal groepen die vanouds vooral met effectiviteitsonderzoek bezig zijn (Louisiana State University, Vanderbilt University).

In een door de International Congress for School Effectiveness and Improvement georganiseerde sessie werden diverse onderzoeken van het London Institute of Education gepresenteerd (Goldstein, Myers, Sammons, Stoll, Thomas, Mortimore en West). In dit instituut bestaat een stabiele traditie op het terrein van effectiviteitsonderzoek. De nadruk ligt er op kwantitatief onderzoek, zowel van min of meer integrale modellen als de aspecten ervan, zoals stabiliteit, contexteffecten en het leren en onderwijzen. Er is echter ook aandacht voor meer op onderwijsverbetering gericht onderzoek, zoals dat van Stoll. Het accent ligt daarbij op de evaluatie van onderwijsverbeteringsprojecten en de feitelijke begeleiding van onderwijsverbetering, maar er wordt ook onderzoek verricht naar bijvoorbeeld de kenmerken van ineffec-

tieve scholen en de wijze waarop ineffectieve scholen kunnen veranderen. In het onderzoek van Myers, Stoll en Reynolds (1996) bleken ineffectieve scholen zich te kenmerken door gebrek aan visie, ongericht leiderschap, disfunctionele onderlinge relaties binnen de staf en ineffectieve praktijken in de klas. Het is echter de vraag of dit niet een vorm van achteraf interpreteren is, in plaats van het via empirisch onderzoek vaststellen van verschillen tussen effectieve en ineffectieve scholen. Bovendien lijkt het erop dat, zoals wel vaker gebeurt, ineffectiviteit wordt gedefinieerd in procestermen in plaats van gerelateerd aan de opbrengsten van onderwijs. Het is een vorm van 'outlier' onderzoek, waarbij scholen met elkaar vergeleken worden op kenmerken waaraan 'effectiviteit' wordt toegeschreven.

Onderwijskundig leiderschap is nog steeds een belangrijk onderwerp in Amerikaans onderzoek op het terrein van onderwijseffectiviteit en er werd uiteenlopend onderzoek op dit terrein gepresenteerd, van de ontwikkeling van een schaal voor 'leader authenticity' en de relatie tussen authenticiteit en effectiviteit - in dit geval de gepercipieerde effectiviteit, die dan ook hoog was (Henderson & Brookhart, 1996) - tot de presentatie van meta-analyses zoals van Hallinger en Heck (1996) over de rol van het schoolhoofd in onderzoek uit de periode 1980-1995. In hun ijver om het belang van onderwijskundig leiderschap empirisch te ondersteunen zijn ze zover gegaan dat een heranalyse van het oorspronkelijke materiaal van Van de Grift (Van de Grift, 1990) werd uitgevoerd om duidelijk te maken dat ook Van de Grift wel de gelijk effecten van onderwijskundig leiderschap had kunnen vinden wanneer hij zijn materiaal op een andere wijze had geanalyseerd.

De presentaties over de resultaten van onderzoek van de integrale modellen voor onderwijseffectiviteit, waarbij diverse niveaus en factoren, variabelen, op de diverse niveaus in beschouwing worden genomen, waren veelal van Europese oorsprong. Het lijkt erop dat daar de interessante ontwikkelingen, zowel wat deelaspecten betreft, zoals stabiliteitsonderzoek, effectgroottes, afzonderlijke componenten, zoals kwaliteitszorg, evaluatie, instructie-effectiviteit, als de evaluatie van onderzoek van integrale modellen vandaan moeten komen en bovendien - voorlopig - uit een vrij beperkt

gebied: Nederland en Engeland. Ander internationaal onderzoek waarover wel in tijdschriften wordt gepubliceerd ontbrak op de AERA.

Het belang van incorporatie van instructie-effectiviteit in integrale effectiviteitsmodellen bleek uit de resultaten van onderzoek van Rowan, Chiang en Miller (1996). De effectgroottes waren weliswaar niet indrukwekkend in deze studie, maar ondersteunen het belang van de effectiviteit van het onderwijzen.

Een interessante sessie gerelateerd aan onderwijseffectiviteit was die over de gelegenheid tot leren. Dit oorspronkelijk binnen de International Association for the Evaluation of Educational Achievement (IEA) geïntroduceerde begrip om de overlap tussen het curriculum en de toets aan te duiden is langzamerhand aardig in betekenis opgerekt. In de Verenigde Staten is er bovendien bijgekomen dat het een belangrijke politieke lading heeft gekregen in verband met een discussie over het kerncurriculum in relatie tot gelijke kansen. 'Opportunity to learn' ging naast toets-curriculumoverlap ook betekenen: 'content coverage', 'content exposure', 'content emphasis' en in de nieuwe TIMSS-studie (Third International Mathematics and Science Study) is het ook de 'quality of instructional delivery'. Daarmee is 'opportunity to learn' het overkoepelende begrip voor instructie-effectiviteit geworden (Schmidt, Valverde, Houang, Wiley, Wolfe & Bianchi, 1996). Behalve tijd ('time on task') lijkt nu ook de kwaliteit van instructie binnen het begrip 'opportunity' gebracht te worden, hetgeen naar mijn mening een te ver gaande oprekking van het begrip is.

Een interessant onderwijsverbeteringsproject werd gepresenteerd door de Memphis 21st Century Schools Exposition, een groep die ook de organisatie van het volgende congres van de International Congress for School Effectiveness and Improvement in 1997 in handen heeft. In Memphis werd de mogelijkheid geboden aan scholen om in diverse vormen van schoolverbetering te participeren. Scholen konden zelf kiezen op welk programma zij intekenden. Er is een uitgebreid scala van programma's voorhanden, van een soort 'outward bound'-achtig programma tot dat van Slavin: 'Roots and Wings', dat sterk gestructureerd en onderwijsgericht is. Tijdens een bijeenkomst werd informatie over alle programma's gegeven en

scholen konden vervolgens intekenen. Op de AERA werden de resultaten van de keuze en de motieven ervoor gepresenteerd; later zal ook vergelijkend onderzoek naar de effecten plaatsvinden. Het probleem dat zich bij deze vergelijkende evaluatie voordoet is dat scholen geen vergelijkbare achtergrond hebben, zowel wat de doelstellingen betreft als wat de leerlingpopulatie en de effecten van het onderwijs betreft. Uit de eerste resultaten van het onderzoek blijkt dat scholen een programma hebben gekozen dat grotendeels aansluit bij de vigerende praktijk in de school voorafgaande aan het project.

Hoewel het gebodene misschien niet altijd meevalt, zoals uit deze korte beschrijving moge blijken, blijft de AERA-conferentie natuurlijk toch de perfecte gelegenheid om een indruk te krijgen hoe onderzoek en praktijkontwikkeling er op bepaalde terreinen bijstaan. In de weinig gestructureerde warwinkel van opvattingen over schoolverbetering min of meer in relatie met onderwijseffectiviteit, was de presentatie van Miles op de 'business meeting' van SIG 'School Effectiveness and School Improvement' een lichtpunt. Miles is zeker geen groot voorstander van effectiviteitsonderzoek, zeker niet in kwantitatieve zin - hij laat zich ook weinig aan kwantitatief onderzoek gelegen liggen - maar binnen het kwalitatieve onderzoek, waarin aanbevelingen gedaan worden voor de verbetering van de onderwijseffectiviteit, gaat hij logisch, analytisch en argumenterend te werk. Hoewel ik er niet aan moet denken dat zijn droom over de stand van zaken in onderwijseffectiviteit in de 21ste eeuw namelijk: alleen maar kwalitatief onderzoek over onderwijs in verandering en een 'theorie' waarin dit wordt vastgelegd, ooit uit zal komen.

Literatuur

- Creemers, B. P. M., & Reezigt, G. J. (in druk). School level conditions affecting the effectiveness of instruction. Geaccepteerd voor publicatie in *School Effectiveness and School Improvement*.
- Grift, W. van de (1990). Educational leadership and academic achievement in elementary education. *School Effectiveness and School Improvement*, 1(1), 26-41.
- Hallinger, P., & Heck, R. H. (1996). *The principal's role in school effectiveness: An assessment of substan-*

- tive findings, 1980-1995*. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, New York.
- Henderson, J. E., & Brookhart S. M. (1996). *Leader authenticity: Key to organizational climate, health and perceived leader effectiveness*. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, New York.
- Myers, K., Stoll, L., & Reynolds, D. (1996). *Understanding ineffectiveness*. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, New York.
- Reynolds, D., Hopkins, D., & Stoll, L. (1993). Linking school effectiveness research: Items for a research agenda. *School Effectiveness and School Improvement*, 4(1), 37-58.
- Rowan, P., Chiang, F.-S., & Miller, R. J. (1996). *Using research on employee performance to study teaching effectiveness: Results from an analysis of teacher effects on student achievement in mathematics using NELS:88 Data*. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, New York.
- Schmidt, W. H., Valverde, G. A., Houang, R. T., Wiley, D. E., Wolfe, R. G., & Bianchi, L. J. (1996). *Studying the intended, implemented and attained curriculum: Strategies for measurement and analysis in large-scale international research*. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, New York.
- Taylor, D. L., Tashakkori, A., & Hardwick, J. (1996). *A typology based on teachers' desire for participation in decision making*. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, New York.

Divisie J: Postsecondary Education

(J. F. M. J. van Hout, Universiteit van Amsterdam)

Ik ga meestal om de paar jaar naar de AERA-conferentie om te bezien of er nieuwe trends zijn waar te nemen in het Amerikaanse onderzoek van het hoger onderwijs. Dit jaar was ik er bovendien om als discussiant op te treden bij de VOR-sessie die door de divisie Hoger Onderwijs werd georganiseerd. Vanuit beide perspectieven geef ik een bijdrage aan deze kroniek.

Ook dit jaar was het onderwerp 'student drop-out, student attrition, student succes' weer een 'evergreen'. De invloed van Tinto, Pascarella en Terenzini is nog steeds groot, mede door een niet aflatende stroom van eigen onderzoek over dit onderwerp. De beleidsrelevantie van dit onderwerp is hoog. In de USA is er een grote uitval van studenten na het eerste studiejaar: bijna 50% bij 2-jarige colleges, ongeveer 25% bij 4-jarige instellingen en 10% bij de zeer selectieve instellingen (Tinto, 1993). Zeker als men bedenkt, dat in de USA bijzonder veel aandacht wordt besteed aan plaatsing en toelating van studenten, zowel van de kant van de instellingen als van de kant van de studenten en hun ouders, is dat een opmerkelijk feit.

Het meest interessante paper over dit onderwerp vond ik dat van Braxton, Shaw Sullivan en Johnson (Vanderbilt University en Belmont University, Nashville). Zij presenteerden een diepgaand review van onderzoeken die het model van Tinto als theoretisch kader hadden gebruikt. Zij testten daarbij de 15 relaties waaruit het model bestaat en relateerden de resultaten aan de theoretische kritiek op het model. Slechts een beperkt deel van de theoretische relaties in het model wordt door het empirisch onderzoek ondersteund. Hierbij blijken tevens verschillen te bestaan tussen typen instellingen en tussen vrouwelijke en mannelijke studenten. De auteurs verwerpen het model als geheel niet, maar zijn wel van mening dat ingrijpende aanpassing van het model noodzakelijk is, waarbij vooral gebruik gemaakt moet worden van andere theoretische invalshoeken, zoals organisatiekunde, economie en psychologie. Zij zijn tevens van mening dat het model voor 'institutional research' een goed uitgangspunt is.

Een tweede onderwerp dat vaak in Divisie J aan de orde komt is het universitaire personeel, 'faculty' of 'administration'. Onderwerpen zijn dan bijv. onderzoeksproductiviteit, satisfactie, onderwijsoriëntatie. Opmerkelijk was dat een aantal sessies hierover georganiseerd was door 'the Committee on the role and status of women in educational R&D'. De beperkte opname van vrouwen en raciale minderheids-groepen in de wetenschappelijke staf is in de USA een groot probleem. De succesvolle Amerikaanse onderzoeker is blank en van het mannelijk geslacht. Dit type onderzoek leek

mij ook voor de Nederlandse universiteiten interessant, omdat Nederland ongeveer de laatste plaats in de westerse wereld inneemt wat het aantal vrouwen in de hogere wetenschappelijke rangen betreft.

Naast enkele 'N=1 studies' was het paper van Bonous-Hammarth (UCLA) een goede uitschieter. Uit haar onderzoek blijkt, dat mannen meer publiceren dan vrouwen en blanken meer dan Afro-Americans. Naast enkele persoonskenmerken (bijv. ambitie), omgevingskenmerken (bijv. takenpakket), blijken zowel ras (als zelfstandige factor) en 'person-environment fit' belangrijke predictoren. De 'P-E fit' is een maat voor overeenkomst in waarden en opvattingen van de respondent met (onderzoekers) de werkomgeving. Zij is ontleend aan onderzoek naar arbeidssatisfactie in andere organisaties en blijkt dus ook relevant voor universiteiten. Vraag is dan wel hoe een 'P-E fit' ontstaat. Dit onderzoek maakte voor mij nog weer eens duidelijk, dat universiteiten niet alleen maar talent en doorzettingsvermogen belonen, maar ook andere criteria en regels hanteren die het voor bepaalde groepen (bijv. vrouwen) moeilijk maken tot de top door te dringen. Bovendien blijkt ook dat problemen die (voorlopig nog) specifiek Amerikaans lijken te zijn, aspecten in zich hebben die in meer algemene zin gelden voor universiteiten en professionele organisaties. De problematiek van raciale minderheden in het wetenschappelijk corps zegt meer over de universiteit zelf, dan ik in eerste instantie had gedacht.

Ieder jaar krijgt de VOR de gelegenheid een sessie op de Annual Meeting van de AERA te vullen met Nederlands onderzoek. Dit jaar was de divisie hoger onderwijs aan de beurt en betrof het Nederlands onderzoek naar studieloopt-banen in het hoger onderwijs. Gepresenteerd werd onderzoek van het IOWO (KU Nijmegen), ROA/TO (RL/UT), SCO/KI (UvA) en SEO (UvA). Gezien de aard van de kroniek zal ik deze bijdragen niet inhoudelijk bespreken. Er zijn twee punten die ik wel vermeldenswaard voor deze kroniek vind. De eerste is, dat we ons over de kwaliteit van dit onderzoek geen zorgen hoeven te maken. Zowel op methodologisch als theoretisch vlak deed het niet onder voor onderzoek dat door Amerikaanse collegae over dit onderwerp werd gepresen-

teerd. Dit raakt meteen mijn tweede punt, namelijk het nut van een dergelijke AERA-sessie die door de VOR jaarlijks gevuld mag worden. De papers zijn niet op de gebruikelijke manier geselecteerd, maar zouden bij een anonieme beoordeling door Amerikaanse peers normaal gesproken ook geaccepteerd zijn. Opname in een paper- of postersessie met vergelijkbare Amerikaanse papers had voor de hand gelegen. In dat geval hadden deze papers en presentatoren meer bekendheid bij onze collegae gekregen dan nu het geval was. De VOR-sessie trok namelijk vooral Nederlandse belangstellenden! Zowel voor de kwaliteit van de papers en presentaties, als voor het opbouwen van netwerken lijkt het mij beter als Nederlandse onderzoekers via de koninklijke weg de Annual Meeting van de AERA binnen stappen, dan via een zij-ingang. De papers van divisie J zijn gemiddeld niet van dat bijzondere niveau, dat Nederlands onderzoek daar uit de toon zou vallen.

Literatuur

- Bomous-Hammarth, M. (1996). *Succeeding in the academy: Predicting the influence of value congruence on the research productivity of black and white scientists*. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, New York.
- Braxton, J. M., Shaw Sullivan, A., & Johnson, R. M. (1996). *Appraisal Tinto's theory of college student departure*. Paper presented at Annual Meeting of the American Educational Research Association, New York City, New York.
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago: University of Chicago Press.

Divisie K: Teaching and Teacher Education

(D. Beijaard en J. Vermunt, ICLON, Rijksuniversiteit Leiden)

Erg veel presentaties binnen Divisie K stonden in het teken van de kennis van leraren en de wijze waarop leraren het best (na)geschoold kunnen worden. Tamelijk veel aandacht werd eveneens besteed aan het beoordelen van leraren (in opleiding). Hieronder wordt vooral op beide eerstgenoemde thema's ingegaan. Enige extra aandacht wordt besteed aan de door de Holmes Group gepresenteerde bevindingen

over de huidige stand van zaken met betrekking tot de opleiding van leraren in de Verenigde Staten.

Het onderzoek naar de kennis van leraren is overwegend kleinschalig en kwalitatief van aard. Het gaat hier vooral om narratief onderzoek of onderzoek waarin het verhaal ('story') van een leraar centraal staat. De argumenten die pleiten voor dit type onderzoek lopen zeer uiteen. Doyle hield bijvoorbeeld een pleidooi voor narratief onderzoek naar aanleiding van door hem gesignaleerde tekortkomingen van het traditionele, empirisch-analytische onderzoeksmodel: 'waarheid' is een communicatieve aangelegenheid en is niet gebonden aan een bepaalde methode. Andere argumenten hebben bijvoorbeeld betrekking op het doorvoeren van veranderingen of vernieuwingen. Elbaz stelde dat het voor het welslagen van veranderingen of vernieuwingen een voorwaarde is om te luisteren naar hetgeen leraren te vertellen hebben over de aan de orde zijnde materie. Zij vindt dat zowel de onderzoeker als de leraar beiden hun voordeel moeten doen met narratief onderzoek. De moeilijkheden van dit soort onderzoek moeten volgens Elbaz niet worden onderschat. Over de functies die een onderzoeker met dit soort onderzoek nastreeft moet duidelijkheid bestaan evenals wiens 'voice' (van de onderzoeker of de onderzochte) uiteindelijk wordt beschreven in het onderzoeksrapport en het methodologische proces op basis waarvan primaire data uiteindelijk worden gerepresenteerd.

Desalniettemin waren er veel presentaties die niet verder gingen dan het beschrijven van unieke situaties met daarin opgenomen veel letterlijke citaten van leraren. Een flink aantal van deze verhaalbeschrijvingen was volstrekt a-theoretisch: zowel een theoretische kader als een terugkoppeling van resultaten naar enige vorm van theorie ontbraken veelal. In een aantal gevallen wordt dit bewust gedaan, bijvoorbeeld vanuit een zeer emancipatoire gedachte (onderzoekers moeten luisteren naar wat de praktijk hun te vertellen heeft) of op basis van een volgestrekte afkeer van theorie. Meestal is echter sprake van onwetendheid, hetgeen als zorgwekkend moet worden aangemerkt. Onderzoekers durven zelfs op basis van een N=1 studie te generaliseren en vinden het vaak niet terecht wanneer ze daarop aangesproken worden.

Inhoudelijk waren de presentaties over de kennis van leraren zeer versnipperd. Interessant was bijvoorbeeld de presentatie van Sherin over het analyseren van de kennis van leraren en het onderscheid tussen 'cognitive modeling' en 'knowledge system analysis'. De eerste verwijst naar het werk van Schoenfeld waarin op basis van een cognitief model fijnmazige en opeenvolgende analyses worden ondernomen van kennis waarover iemand beschikt, de tweede verwijst naar het werk van Shulman en Grossmann die vooral zijn geïnteresseerd in de genese en ontwikkeling van kennis. Ook wees Sherin erop dat het goed is te onderscheiden tussen vorm (structuur en organisatie) en inhoud (waar iets over gaat en waartoe het dient) van kennis. Het meeste onderzoek naar de kennis van leraren is vooral gericht op de vorm ervan.

Verhalen ('stories') van leraren worden vaak gebruikt als een middel voor scholing van leraren (in opleiding). In diverse presentaties werd duidelijk dat zij reflectie bij leraren bevorderen. Andere middelen voor scholingsdoeleinden waarvan een positieve invloed uitgaat op de eigen professionele ontwikkeling zijn bijvoorbeeld conferentie-achtige bijeenkomsten. Tijdens zulke bijeenkomsten, die overigens goed in het totale curriculum moeten worden gepland, worden perspectieven van anderen (peers) ingezet om de complexiteit van de eigen onderwijssituatie beter te begrijpen. Daarnaast werd er tijdens diverse presentaties op gewezen dat (a.s.) leraren ook veel kunnen leren van hun eigen biografie.

In nogal veel presentaties werd de nadruk gelegd op de rol van de praktijk bij de opleiding van leraren. De achtergrondgedachte daarbij is dat ervaring de beste leermeester is. Chin en Russell presenteerden bijvoorbeeld het opleidingsmodel van de universiteit waarbij zij werkzaam zijn. Dit model gaat sterk uit van praktijkleren: na een korte introductie verkeren de studenten lange tijd in de praktijk. De kwaliteit van hun leerproces is daarmee sterk afhankelijk van de begeleiding van leraren in de stagescholen. Op deze en andere modellen waarin een eenzijdig accent op de praktijk ligt, werd flink kritiek geleverd. Bullough, bijvoorbeeld, verwijst zo'n opleiding dan dat die te veel is gebaseerd op een 'craft' model. Leren van ervaring zonder daarbij een perspectief te krijgen

is in zijn ogen onjuist, waarbij onder andere het gevaar bestaat dat leraren in opleiding misconcepties ontwikkelen. Bovendien is zo'n opleiding ethisch onjuist: kinderen kunnen niet zomaar aan 'blanco' leraren worden overgelaten.

In andere presentaties werd te sterk van de praktijk afhankelijke lerarenopleidingen onder meer verweten a-theoretisch te zijn. In dit verband moet worden gewezen op het onderzoek van Nedell. Uit dit onderzoek blijkt bijvoorbeeld dat er een positieve samenhang is tussen professionele ontwikkeling en het beschikken over kaders ('frameworks') om de eigen praktijk te analyseren. Door opleiding verworven kennis kan wel degelijk een belangrijke rol spelen in het proces van leren onderwijzen. Het lijkt erop dat deze opvatting weer enige terreinwinst boekt.

Verandering in de lerarenopleiding werd vooral bepleit naar aanleiding van een door de Holmes Group gehouden onderzoek over ontwikkelingen die zich de laatste tien jaar hebben voltrokken in de lerarenopleidingen in de Verenigde Staten en waar het de komende jaren met deze opleidingen naartoe moet. Er werd vooruit gelopen op het nog in 1996 te publiceren, derde rapport van deze groep, getiteld 'Tomorrow's schools of education'. (Het eerste rapport, 'Tomorrow's teachers', verscheen in 1985; het tweede werd in 1990 uitgebracht onder de titel 'Tomorrow's schools'). Alle lerarenopleidingen die lid zijn van de Holmes Group (N=190) evenals een aantal personen zijn op een aantal voor lerarenopleidingen relevante punten onderzocht of bevraagd, zoals het streven naar goede prestaties, handelingsplannen, onderzoek in de lerarenopleiding, organisatiegraad en leiderschap. Op basis van de onderzoeksresultaten moet worden geconcludeerd dat veranderingen in de lerarenopleidingen nog steeds in de kinderschoenen staan. Gepleit wordt voor een meer omvattend plan van aanpak dan in 1985 werd voorgesteld in het rapport 'Tomorrow's teachers'. In de nabije toekomst moet op basis van dit plan gewerkt worden aan het ontwikkelen van het curriculum voor de lerarenopleiding, het formuleren van voorwaarden om continue professionele ontwikkeling van leraren te realiseren, het ontwikkelen van onderzoeksprogramma's c.q. -agenda's en het houden van debatten over wat een lerarenopleiding nu eigenlijk precies moet zijn.

Diverse bekende wetenschappers was gevraagd te reageren op hetgeen door de onderzoekers van de Holmes Group werd gepresenteerd. Kelley onderkende het algemene effect dat van het werk van de Holmes Group uitgaat. Hij stelde dat niemand is geïnteresseerd in het programma van de lerarenopleiding; het gaat om de verbinding van dit programma met hetgeen in scholen gebeurt. In dit opzicht miste hij enige aandacht bij de Holmes Group voor de evaluatie van leraren via leerresultaten van leerlingen. Lanier ging in haar reactie in op het gegeven dat lerarenopleidingen niet altijd onderzoek doen. Zij acht onderzoek van groot belang voor het ontwikkelen van een kwalitatief goede lerarenopleiding. Hargreaves was in zijn reactie het meest uitgebreid en sprak vooral vanuit de positie van onderzoeker die kritiek levert op een onderzoek. Het doel van een lerarenopleiding is in zijn ogen niet zozeer een goede voorbereiding op het leraarsberoep op zich, maar moet vooral gericht zijn op het verbeteren en veranderen in de scholen zelf. Ook hekelde hij in dit verband de samenwerking van lerarenopleidingen met de zogenoemde 'professionele ontwikkelingsscholen': leraren leren onderwijzen in deze scholen, maar komen daarin na hun afstuderen niet te werken. Het zou de Holmes Group volgens hem niet al-

leen te doen moeten zijn om de status van leraren te verhogen, maar ook om het ontwikkelen van hoge standaarden waaraan leraren zouden moeten voldoen.

Binnen de divisie onderwijzen en lerarenopleiding, tenslotte, was het moeilijk onderzoek te vinden over kennisconstructieprocessen van leraren-in-opleiding. Wel waren er diverse presentaties over verschillende opleidingsdidactische modellen en de veronderstelde effecten ervan. Deze hadden over het algemeen echter weinig oog voor de wijze waarop aanstaande leraren de didactische maatregelen die deel uitmaken van deze modellen interpreteren en gebruiken. Ook de strategieën die zij hanteren om hun kennisbasis op te bouwen en bijvoorbeeld de kloof tussen hun ervaringskennis en theoretische kennis te overbruggen, bleven sterk onderbelicht.

Aan deze kroniek werkten mee: *D. Beijaard, E. De Corte, B. P. M. Creemers, F. Dochy, J. Elen, J. F. M. J. van Hout, E. Knoors, J. Lowyck, G. W. Meijnen, M. Mulder, N. Verloop, J. Vermunt en L. Verschaffel*

De eindredactie werd verzorgd door *N. Verloop en L. Verschaffel*