

Formatieve curriculumevaluatie en onderwijspsychologie

N. VERLOOP*

Nijmeegs Instituut voor Onderwijsresearch, K.U. Nijmegen

Samenvatting

In dit artikel wordt een formatief evaluatie-onderzoek beschreven van een curriculum dat in sterke mate gebaseerd is op een aantal leerpsychologische uitgangspunten. Het betreft een leergang rekenen voor klas 2 van het basisonderwijs die is ontwikkeld door het Schooladviescentrum te Utrecht en momenteel door ongeveer 400 scholen wordt gebruikt.

Formatieve curriculumevaluatie wordt hier gezien als een poging op een zo verantwoord mogelijke wijze te komen tot waarderende uitspraken over een curriculum, waarop revisie-beslissingen kunnen worden gebaseerd.

Omdat het vertrekken vanuit een aantal leerpsychologische principes als het meest karakteristiek voor dit curriculum is te beschouwen, worden in dit artikel, na een algemeen overzicht van het onderzoek, alleen die gedeeltes van het onderzoek uitgebreid besproken waarin deze leerpsychologische component een centrale rol speelt: het doelstellingen-gedeelte en het observatie-gedeelte.

In het doelstellingen-gedeelte werd bij 42 leerkrachten die met het curriculum werkten nagegaan wat hun mening was over een groot aantal aan het curriculum ten grondslag liggende doelstellingen. Tevens werd getracht wat meer greep te krijgen op de impliciete structuur waarvan de leerkrachten uitgaan bij hun denken over het reken-curriculum.

In het observatie-gedeelte werd vervolgens een steekproef van 14 leerkrachten gedurende bijna drie uur tijdens het onderwijsleerproces geobserveerd. Daarbij werd o.m. met behulp van een observatie-instrument vastgesteld in welke mate de leerpsychologische uitgangspunten van het curriculum doorwerkten in de concrete klassensituatie. De relatieve betekenis van de diverse leerpsychologische principes voor de uiteindelijke leerling-effecten werd nagegaan.

Informatie uit de afzonderlijke evaluatie-instrumenten en combineren van de gegevens leidde op een groot aantal punten tot conclusies en aanbe-

velingen ter herziening van het curriculum en de begeleiding.

1. Inleiding. Enkele theoretische achtergronden**

1.1. Onderwijspsychologie en onderwijspraktijk

Vele onderwijskundigen zijn de laatste jaren overtuigd geraakt van het feit dat de onderwijspsychologie van veel groter betekenis zou kunnen zijn voor de dagelijkse gang van zaken in de school dan thans het geval is. Tegelijkertijd is vanuit de onderwijspsychologie een steeds sterker wordende tendens merkbaar tot 'vertaling' en 'toepassing' van verworven inzichten, evenals een algemene gerichtheid op die vormen van leren en informatieverwerking, die ook voor het schoolse leren relevant zijn.

De functie die de onderwijspsychologie voor de dagelijkse schoolpraktijk kan hebben is waarschijnlijk samen te vatten in de volgende twee categorieën:

- a. Het toerusten van (a.s.) leerkrachten met kennis en vaardigheden op onderwijspsychologisch gebied, op zodanige wijze dat zij daarvan gebruik kunnen maken in hun lesvoorbereiding en hun lesgedrag. Aan deze toepassingsmogelijkheid is tot nu toe weinig aandacht besteed.
- b. Het ontwerpen en structureren van curricula (in de zin van: planningsdocumenten voor reeksen van onderwijsleersituaties) op zodanige wijze dat daarin rekening wordt gehouden met bepaalde relevante onderwijspsychologische inzichten en principes. Momenteel worden in toenemende mate pogingen ondernomen deze tweede toepassingsmogelijkheid te realiseren, met name voor de sekwentering, zowel op leergangniveau (sekwentering van leerstofinhoud) als op lesniveau (lesopbouw).

* Per 1 maart 1979: Centraal Instituut voor Toetsontwikkeling.

** Met dank aan Pieter Appelfhof en Michel Zwarts voor hun opbouwend commentaar op het concept van dit artikel.

Elders is aangegeven (Verloop 1978) dat voor een dergelijke functie waarschijnlijk slechts die onderwijspsychologische theorieën in aanmerking komen die zich op de procesmatige aspecten van het leren richten (i.p.v. op de produkt-aspecten), waarbij deze theorieën voorschriften kunnen geven voor de wijze waarop een dergelijk leerproces kan worden opgebouwd. Een tweetal groepen theorieën bleek met name hiervoor geschikt te zijn, nl. (1) een groep uit de VS afkomstige informatieverwerkingstheorieën (Klahr 1976, Anderson, Spiro en Montague 1977) evenals het model van Gagné en Briggs (1974), en (2) een aantal uit de Sovjet-Unie afkomstige theorieën, (Cole en Malzman 1969, Van Parreren en Carpay 1972). Tevens werd aangegeven wat de perspectieven zijn voor een 'theorie van het onderwijzen' wanneer men gebruik maakt van dergelijke onderwijspsychologische theorieën.

We zullen ons in dit artikel voornamelijk richten op de curriculumproblematiek, speciaal de evaluatie van een dergelijk curriculum. Wanneer men immers in de opbouw van een curriculum heeft getracht gebruik te maken van bepaalde onderwijspsychologische inzichten en principes en men vervolgens d.m.v. formatieve evaluatie verbeteringen in dit curriculum wil aanbrengen, zullen deze onderwijspsychologische principes een centraal element in de evaluatie moeten vormen. Dit impliceert ook dat bij de evaluatie de nadruk zal liggen op hetgeen zich op cognitief gebied procesmatig in de klassensituatie afspeelt. Het evaluatie-onderzoek dat hier wordt gerapporteerd is dus een illustratie van een curriculumevaluatie waarin een aantal onderwijspsychologische principes centraal staan. Met de toenemende belangstelling voor onderwijspsychologische (sekwenterings)principes bij de opbouw van curricula zullen zij in de toekomst waarschijnlijk een steeds grotere rol gaan spelen bij curriculumevaluatie.

1.2. *Formatieve curriculumevaluatie*

Een van de belangrijkste gedeelten van het hier gerapporteerde evaluatie-onderzoek is dat waarin aan de hand van het gedrag van de leerkracht wordt nagegaan of de leerkracht het curriculum uitvoert overeenkomstig de bedoelingen van de ontwikkelaars. Dit sluit aan bij de door Creemers gesignaleerde tendens een sterker verband te leggen tussen enerzijds het onderzoek van het leerkrachtgedrag en anderzijds het onderzoek naar curriculumontwikkeling, met name curriculumimplementatie, (Creemers 1978, p. 7).

In een aantal bronnen (bijv. Worthen en Sanders 1973) wordt uitvoerig ingegaan op datgene wat in het

algemeen karakteristiek zou zijn voor evaluatie-onderzoek, vooral in vergelijking met het klassiek experimenteel onderzoek. Een karakteristiek die door de meeste auteurs wordt aangegeven en die ook voor het onderhavige onderzoek opgaat is het 'beslissingsgerichte' karakter van evaluatie-onderzoek (Popham 1975, Worthen en Sanders 1973, Stufflebeam 1974, Edwards 1974). Bij evaluatie-onderzoek gaat het erom gegevens te leveren die van belang zijn voor het nemen van concrete beslissingen. Met betrekking tot dit onderzoek hadden de beslissingen te maken met de revisie van het curriculum. De evaluatie is daarom formatief van karakter, (vgl. Scriven 1967). De begrippen formatief-summatief worden hier overigens niet opgevat als een tegenstelling (zoals bij Scriven), maar als uiteinden van een continuüm (vgl. Popham 1975, of Meijer 1978).

Een vaak terugkerende controversie wordt gevormd door de vraag of bij (formatieve) evaluatie gebruik kan worden gemaakt van klassieke experimentele onderzoeksdesigns. Aan het ene uiterste staan hier degenen die met het toepassen van deze designs in evaluatie-onderzoek geen enkele moeite hebben, zoals Wolf (1974) die bij het opzetten van een evaluatie-onderzoek vanuit de data-analyse lijkt te vertrekken, of Glass (1973) die bijv. bij de introductie van zijn 'intact group design' terloops opmerkt dat bij het evaluatie-onderzoek hier alle leerlingen random aan de klassen moeten worden toegewezen. Aan het andere uiterste staan degenen die in zo sterke mate recht willen doen aan de eigen aard van het evaluatie-object dat ze de empirisch-analytische werkwijze als zodanig afwijzen en uitsluitend gebruik willen maken van kwalitatieve methoden. De evaluator heeft hier als voornaamste taak het samen met de direct betrokkenen verhelderen van hun eigen probleemsituatie ('responsive evaluation', Scheyer en Stake 1976), eventueel gebruikmakend van bijv. de methoden der kunstcritiek ('educational connoisseurship', Eisner 1976) of journalistiek ('portrayal evaluation', MacDonald 1976). In het hier gerapporteerde onderzoek is de ontoereikendheid van de klassieke experimentele designs voor tal van evaluatie-vragen onderkend. Er is getracht daarvoor een oplossing te vinden door per evaluatie-object een aantal 'meet-punten' te creëren (vgl. par. 3).

Hoewel in dit onderzoek wel getracht is een aantal variabelen te isoleren die een bijdrage leveren aan het eind-effect, wordt hier niet de door Cronbach gestelde eis onderschreven dat in elk evaluatie-onderzoek zou moeten blijken 'how the course produces its effects', (vergelijk voor een tegengestelde opvatting Scriven 1967, Stufflebeam 1974).

Ten slotte, in dit onderzoek is er vanuit gegaan dat

het de taak van de evaluator is op grond van zijn gegevens aanbevelingen, en daarmee waarde-oordelen te formuleren ten aanzien van het object van evaluatie, terwijl vervolgens de beslissing aan de beleidsbepalers (i.c. het ontwikkelingsteam) wordt overgelaten. Het is bekend dat, hoewel evaluatie oorspronkelijk door bijv. Cronbach en Scriven op deze wijze is opgevat en ook nu nog vaak op deze wijze wordt gedefinieerd (bijv. MacKay en Maguire 1971, p. 4), er een belangrijke stroming is die deze rol nadrukkelijk aan de evaluator ontzegt, (Alkin, Provo, Stufflebeam, Guba).

Voor een meer gedetailleerde bespreking en motivering van de in dit onderzoek gedane theoretische en methodologische keuzen wordt verwezen naar het oorspronkelijke onderzoeksverslag, (Verloop 1977).

2. Object van evaluatie

Het curriculum waarom het in dit onderzoek gaat omvat een volledige rekenleergang voor klas 2 van het basisonderwijs; zij vormt een onderdeel van een leergang rekenen voor klas 1 t/m 4 en is ontwikkeld door het Schooladviescentrum (SAC) te Utrecht, (Nelissen, Klukhuhn en Vuurmans 1978). De eerste ontwikkelingswerkzaamheden dateren uit 1972 en zijn ontstaan vanuit een bij de leerkrachten geconstateerde ontevredenheid over bestaande rekenmethodes. Het curriculum heeft zich in een snel tempo verspreid via andere schoolbegeleidingsdiensten. Momenteel wordt op ongeveer 400 basisscholen in Nederland met het curriculum gewerkt. De ontwikkelingsgroep werkt aan een uitbreiding van het curriculum tot de bovenbouw.

De elementen die bij het evaluatie-onderzoek zijn betrokken zijn vastgesteld in overleg met de ontwikkelingsgroep, waarbij vooral werd gelet op datgene wat als typerend voor dit reken-curriculum beschouwd kon worden, ('onderscheidende kenmerken', vgl. Creemers en Meijer 1978, p. 12 e.v.). De belangrijkste karakteristiek van dit SAC-rekenprogramma is dat men sterk de nadruk legt op inhoudelijke leerstofvernieuwing van het rekenonderwijs, gekoppeld aan een zeer expliciete opvatting over de wijze waarop het leerproces bij de leerling zou moeten verlopen, terwijl men allerlei organisatorische zaken m.b.t. de inrichting van het onderwijsleerproces als secundair beschouwt. De ontwikkelaars zijn van mening dat de meeste in gebruik zijnde rekenmethodes met hun zeer uitgebreide aandacht voor het uitwerken van een model voor de organisatie van de differentiatie, voorbijgaan aan de essentie

van het rekenonderwijs: óf men verzuimt aandacht te geven aan de begrippen en procedures die essentieel zijn voor het mathematisch leren denken óf men gaat volkomen voorbij aan de leerpsychologische weg waarlangs het leren van het kind zich zou moeten voltrekken óf men laat het op beide punten afweten. Los van de, hier niet aan de orde zijnde, vraag of men zich met een dergelijk standpunt kan verenigen kan geconstateerd worden dat een dergelijk uitgangspunt verstrekkinge gevolgen heeft voor de vormgeving van het curriculum. Voor het identificeren van de mathematische kernbegrippen die essentieel zijn voor een werkelijk inzicht in allerlei rekenkundige/wiskundige relaties baseren zij zich op het oordeel van deskundigen in het wiskundig vakgebied; vooral vanuit de school van de Russische leerpsycholoog Davydov is op deze kernbegrippen de nadruk gelegd. Voor het bepalen van de leerpsychologische weg die het kind moet volgen om tot abstracte kennis van een begrip of een relatie te komen maakt men vooral gebruik van de ideeën van de Rus Gal'perin. Met name de wijze waarop het kind zich volgens deze theorie via een interiorisatieproces beweegt van het concrete manipuleren met materiaal naar een op volkomen abstract niveau verlopende handeling is duidelijk in het programma terug te vinden. Omdat de begrippen uit deze theorie voor een belangrijk deel de basis voor de onderwijspsychologische opbouw van het curriculum vormen is, in overleg met de ontwikkelingsgroep, deze theorie als kern van de onderwijspsychologische component van de evaluatie genomen. Voor een gedetailleerde bespreking van deze theorie wordt verwezen naar Gal'perin (1969) of Van Parreren en Carpay (1972). Volgens deze theorie voltrekt de genoemde interiorisatie zich via een aantal stadia, waarbij een aantal 'niveaus' van de handeling worden doorlopen; in het begin wordt de handeling concreet, op 'materieel' niveau voltrokken, daarna volgt de 'perceptieve' handeling waarbij als het ware met de ogen met het materiaal gemanipuleerd wordt, vervolgens de 'verbale' fase waarbij de steun van het hardop gesproken woord in de plaats van de visuele steun komt en ten slotte volgt, via de fase van het spreken 'voor zichzelf' (door de ontwikkelingsgroep aangeduid als 'handelen met de ogen dicht') de fase waarin de handeling op volkomen abstract niveau wordt voltrokken, de 'mentale' handeling. Bij het aanleren van een handeling verandert deze handeling niet alleen op de parameter 'niveau' (van materieel naar mentaal), maar eveneens op een drietal andere parameters, nl. 'uitvoerigheid', betrekking hebbend op de mate waarin alle deel-componenten van de handeling ook daadwerkelijk worden uitge-

voerd, 'generalisatie', betrekking hebbend op de mate waarin de handeling op specifieke objecten is gericht, en 'beheersing', betrekking hebbend op de mate waarin de handeling als het ware één vloeiende beweging is. Hoewel deze parameters in eerste instantie de karakteristieken zijn van de handeling die door de leerlingen wordt voltrokken kan men ze ook beschouwen als dié handelingskarakteristieken waarop de leerkracht door zijn onderwijs op gerichte wijze invloed moet gaan uitoefenen. Door in zijn onderwijs op een gerichte manier wijzigingen op deze parameters te bewerkstelligen brengt hij de leerlingen tot volwaardige mentale handelingen.

De hier genoemde theoretische constructen vormen het belangrijkste deel van de onderwijspsychologische component van het evaluatie-onderzoek en eveneens een belangrijk deel van het evaluatie-onderzoek als geheel. Omdat in dit artikel vnl. gepoogd wordt een uitwerking te geven van een evaluatie van een curriculum dat gebaseerd is op onderwijspsychologische principes zullen deze in de rest van het artikel voortdurend in het centrum van de belangstelling worden geplaatst. In het reken-curriculum zijn deze principes uiteraard betrokken op bepaalde inhoud. De inhoudelijke indeling van het rekenprogramma zal dan ook voortdurend in het onderzoek terugkeren; daarbij wordt in het kader van dit artikel een inperking gemaakt tot een drietal inhoudscategorieën die op een aantal belangrijke punten onderling vergelijkbaar zijn en tezamen verreweg het grootste deel van het curriculum bestrijken: optellen/afrekken, vermenigvuldigen en 'open bewerkingen' (d.w.z. opgaven waarbij de leerlingen zelf een 'onbekende' op een openstaande plaats moeten uitrekenen). Naast de leerinhoud van het curriculum werden ook diverse in het programma verwerkte differentiatievormen bij de evaluatie betrokken, evenals de bij het programma behorende criteriumtoetsen. Ten slotte vormde ook de door de ontwikkelingsgroep verzorgde schoolbegeleiding (in de vorm van schoolbezoeken en bijeenkomsten met groepen leerkrachten die met het curriculum werkten) voorwerp van evaluatie. We zullen ons hierna echter uitsluitend richten op de onderwijspsychologische component in het curriculum.

3. Globale opzet van het onderzoek

Een bespreking van de opzet van het onderzoek komt in feite neer op het aangeven van de 'meetpunten' waarop de evaluatie-objecten voorwerp van onderzoek waren (vgl. par. 1.2). De evaluatie-objecten zijn aangegeven in paragraaf 2. Ten aanzien van

elk van de evaluatie-objecten werden een aantal evaluatievragen geformuleerd en werden een aantal instrumenten ontworpen voor het verzamelen van gegevens ter beantwoording van deze vragen; ten slotte werden de gegevens van de diverse instrumenten (en van de diverse evaluatie-objecten) met elkaar in verband gebracht. In grote lijnen gaat het hier om de volgende evaluatie-vragen:

- (a) Wat is het oordeel van de leerkrachten over de belangrijkheid van het curriculum-onderdeel.
- (b) Hoe wordt, blijkend uit de observatie-gegevens, dit onderdeel in de concrete klassesituatie gerealiseerd, (zowel wat de leerpsychologische component betreft als meer inhoudelijk).
- (c) Wat is de mening van de leerkrachten omtrent de realiseerbaarheid ervan, i.c. welke problemen heeft men daarbij ontmoet.
- (d) In welke mate komt het onderdeel in de begeleidingsbijeenkomsten aan de orde.
- (e) Wat zijn de toetsprestaties van de leerlingen op dit onderdeel.

Al deze informatie diende bij te dragen tot het nemen van meer verantwoorde beslissingen omtrent revisie van het curriculum.

In de eerste fase van het onderzoek werd getracht samen met de ontwikkelingsgroep, d.m.v. een cyclisch proces van steeds verdere concretisering, te komen tot een inventarisatie van de doelstellingen die aan het tweedeklas-rekenprogramma ten grondslag liggen. Op deze wijze werden een aantal doelstellingen-uitspraken geformuleerd, die zowel betrekking hadden op de inhoudelijke rekenleerstof als op de wijze waarop het aanleren van deze leerstof leerpsychologisch gezien zou moeten verlopen. Deze lijst van doelstellingen werd uitgebreid met een aantal rekendoelstellingen die nadrukkelijk niet door de ontwikkelingsgroep kon worden onderschreven, (zie verder par. 4). De totale lijst van 97 doelstellingen-uitspraken werd ter beoordeling op 'belangrijkheid' (9-puntsschaal) voorgelegd aan alle Utrechtse en Rotterdamse leerkrachten die op dat moment met het tweedeklas-rekenprogramma werkten ($n = 51$). De respons was 83%. Ter beantwoording van de vragen omtrent de realiseerbaarheid van het curriculum in de concrete klassesituatie en van een aantal vragen omtrent toetsgebruik, ontvangen begeleiding etc. werd gekozen voor een (half gestructureerd) interview. Daarbij werd ook ten aanzien van alle curriculum-componenten de vraag gesteld in welke mate men bij de realisering van deze curriculum-component trachtte te werk te gaan volgens de aanwijzingen van het curriculum.

Het werd essentieel geacht gegevens omtrent het functioneren van het curriculum in de concrete klassesituatie ook door middel van eigen waarneming te verzamelen. Daarom werd een steekproef van 14 leerkrachten vervolgens gedurende bijna drie uur tijdens het onderwijsleerproces geobserveerd, (verdeeld over een viertal observaties per leerkracht). Daarbij werd van een tweetal voor dit doel ontwikkelde observatie-instrumenten gebruik gemaakt. Bij het eerste instrument werd nagegaan in welke mate en op welke wijze de leerkracht gebruik maakte van de aan dit curriculum ten grondslag liggende leerpsychologische principes. Bij het tweede instrument ging het om het gebruik door de leerkracht van 'essentiële inhouden' bij de diverse onderdelen, (optellen/afrekken, vermenigvuldigen, open beweringen, verhoudingen, kansberekening, meten). Met 'het gebruiken van essentiële inhouden' wordt hier bedoeld dat bij een bepaald onderdeel, bijv. het vermenigvuldigen, zaken aan de orde moeten komen die essentieel zijn voor het verwerven van inzicht in 'het vermenigvuldigen'. Informatie omtrent hetgeen op de begeleidingsbijeenkomsten aan de orde kwam werd verkregen d.m.v. een inhoudsanalyse van de verslagen van deze bijeenkomsten. De leerlingen-prestaties op de diverse onderdelen werden vastgesteld door middel van reeds bestaande en bij het rekenprogramma behorende criterium-georiënteerde toetsen.

Wanneer men nu de genoemde 'meetpunten' en de evaluatie-objecten tegen elkaar uitzet, ontstaat het overzicht van het onderzoek zoals weergegeven in Figuur 1. Daarbij wordt met een 'x' onder een bepaald instrument aangegeven dat met dit instrument informatie over het betreffende evaluatie-object kan worden verkregen terwijl 'xx' aangeeft dat dit in sterke mate het geval is.

De twee volgende paragrafen hebben betrekking op die gedeelten van het onderzoek die met betrekking tot de onderwijspsychologische component het meest interessant zijn, nl. het doelstellingen-gedeelte en het observatie-gedeelte.

Het *doelstellingen-gedeelte* is vooral van belang omdat daarin o.m. een poging wordt gedaan d.m.v. cluster-analyse wat meer greep te krijgen op de impliciete structuur waarvan de leerkrachten uitgaan bij hun denken over het reken-curriculum, vooral met betrekking tot de inhoudelijke resp. onderwijspsychologische component van het curriculum.

Het *observatie-gedeelte* van de onderwijspsychologische component is van belang omdat daarin een procedure werd ontwikkeld om de doorwerking van de onderwijspsychologische principes in de concrete klassesituatie te registreren en omdat daarin d.m.v. multiple regressie-analyse de relatieve betekenis van de diverse onderwijspsychologische principes voor de uiteindelijke effecten werd nagegaan.

	oordeel leerkrachten over doelen (belangrijkheid)	onderwijsleerproces				condities
		oordeel leerkr.	observatie-gegevens	leerlingen-prestaties	begeleidings-bijeenkomsten	
instrument:	9-punt-schaal	interview/vragenlijst	observ.-instrumenten	toets	inh.-analyse	interview/vragenlijst
evaluatie-objecten:						
optellen/afrekken	XX	X	XX	X	X	
vermenigvuldigen	XX	X	XX	X	X	
open beweringen	XX	X	XX	X	X	
verhoudingen	XX	X	X	X	X	
kansberekening	XX	X	X		X	
meten	XX	X	X		X	
leerpsych.	XX	X	XX		X	
differentiatie	X	XX			X	XX
toetsen	X	XX			X	XX
begeleiding	X					XX

Figuur 1 Totaal-overzicht van de onderzoeksopzet

4. Doelstellingen-gedeelte

4.1. Algemene functie van dit onderzoeksgedeelte

Voor het opnemen van een doelstellingen-gedeelte in dit evaluatie-onderzoek zijn een aantal redenen te formuleren. Ten eerste bleek uit gesprekken met de ontwikkelingsgroep en uit een analyse van het ontwikkelde curriculum dat de ontwikkelaars een aantal goed formuleerbare doelstellingen voor ogen stonden; een meer gedetailleerde formulering en explicitering daarvan zou ook voor de ontwikkelaars verhelderend kunnen werken ten opzichte van hun eigen curriculum. Vervolgens zou op deze wijze nagegaan kunnen worden of de door de ontwikkelaars nagestreefde doelen ook door de betrokken leerkrachten als belangrijk worden beschouwd. Ten slotte zou een gedifferentieerd oordeel van de betrokken leerkrachten over de belangrijkheid van de doelstellingen bij de diverse onderdelen waarschijnlijk meer helderheid kunnen brengen over de vraag waarom bepaalde onderdelen van het curriculum beter functioneren dan andere.

Omdat in het kader van dit onderzoek van de ontwikkelaars redelijkerwijs niet verwacht mocht worden dat zij erg veel tijd zouden investeren in het gedetailleerd specificeren van de doelen voor de diverse onderdelen is op grond van een analyse van het ontwikkelde curriculum een gedetailleerde lijst van doelstellingen samengesteld. Samen met de ontwikkelaars is deze lijst enkele malen besproken en herzien. Bij de uiteindelijke versie was men van mening dat er inderdaad een redelijk beeld van de door de ontwikkelaars nagestreefde doelen werd gegeven.¹ Voor elk van de inhoudelijke onderdelen (optellen/afrekken etc.) werden, naast de curriculum-doelen ('SAC-doelen') ook enkele doelstellingen opgenomen die door de ontwikkelingsgroep niet als belangrijk werden onderschreven ('N-doelen'). Het doel hiervan was het maken van een vergelijking tussen het oordeel van de leerkrachten omtrent deze doelen en de doelen die wél door de ontwikkelaars als belangrijk werden gekenschetst. Een tweede vergelijkingmogelijkheid lag hier uiteraard tussen ontwikkelaars en leerkrachten.²

Op deze wijze werd voor elk onderdeel een groepje doelen geformuleerd. Bijvoorbeeld voor het onderdeel 'vermenigvuldigen' resulteerde dit in een tiental doelstellingen-uitspraken, zoals: 'De leerling moet op de getallenlijn sprongetjes (van bijv. 4) kunnen maken, analoog aan het opzeggen van een tafel (tafel van 4)'. Aan de 42 leerkrachten werd een dergelijke uitspraak voorgelegd en werd gevraagd de belangrijkheid van deze doelstelling aan te geven op

een 9-puntsschaal. Het gemiddelde van deze tien doelstellingen-uitspraken werd opgevat als een score voor de 'belangrijkheid' van het onderdeel 'vermenigvuldigen' volgens deze leerkracht. Met betrekking tot de N-doelen van elk onderdeel werd een zelfde procedure gevolgd. Ook door de 4 ontwikkelaars werden de doelen op deze wijze beoordeeld. Hierdoor werd het mogelijk op zeer gedetailleerde wijze vergelijkingen te maken tussen enerzijds leerkrachten en anderzijds ontwikkelaars omtrent het belang dat aan bepaalde curriculumdoelen werd gehecht.

Wat betreft de vorm van de doelstellingen-uitspraken, er zijn ook een aantal uitspraken opgenomen die niet in gedragstermen geformuleerd zijn. Het belangrijkste criterium voor opname van een doelstelling in de lijst was niet het voldoen aan bijv. Mageriaanse voorschriften maar was of het beoordelen van een dergelijke doelstelling informatie zou kunnen opleveren voor de onderzoeksvraagstellingen. Het is overigens waarschijnlijk dat veel (te recht) tegen gedragsdoelstellingen ingebrachte bezwaren³ voor een 'concreet' vakgebied als het rekenen in veel mindere mate van toepassing zullen zijn. In de hier ontworpen doelstellingslijst komen ook doelstellingen voor die ontleend zijn aan het leerpsychologisch model. Volgens dit model moet de leerling op een bepaald punt in de opbouw van de handeling deze handeling op een expliciet aangegeven wijze uitvoeren; de doelstellingen-uitspraken die hierop betrekking hebben zijn in de lijst opgenomen o.m. om na te gaan in hoeverre de leerkrachten deze leerpsychologische opbouw van belang achten. Zo veel mogelijk werd daarbij het gebruik van specifieke leerpsychologische termen (materiële handeling e.d.) vermeden en vervangen door een omschrijving van het begrip in meer alledaagse termen.

De N-doelen werden, behalve aan de gangbare rekenmethodes, vnl. ontleend aan de bestaande praktijk van het rekenonderwijs, (bijv. het feit dat vaak erg belangrijk vindt leidde tot het opnemen van deze doelstelling als 'N-doel'). Waar dit noodzakelijk leek werden de doelstellingen-uitspraken toegelicht met een zo neutraal mogelijk voorbeeld.

Om te voorkomen dat men bij elkaar behorende doelstellingen als zodanig zou herkennen en daarom op analoge wijze zou beantwoorden, en om bovendien te voorkomen dat groepjes van 'N-doelen' als zodanig herkend zouden worden zijn de uitspraken in random volgorde geplaatst. Eventuele effecten van vermoeidheidsverschijnselen werden opgevangen door verschillende versies in de totaalvolgorde aan te brengen.

4.2. Clustering van doelen

Naast de hierboven besproken functies van het doelstellingen-gedeelte in dit onderzoek werd van een deel van de hier verkregen data ook nog een anderssoortig gebruik gemaakt. In het kader van dit evaluatie-onderzoek werd het interessant geacht na te gaan of de leerkrachten bij de beoordeling van de curriculum-doelstellingen een refentiekader hanteerden dat vnl. inhoudelijk óf vnl. leerpsychologisch van karakter was. In dit onderzoek werd een eerste exploratieve poging gedaan deze vraag te beantwoorden. Daarbij werd als volgt te werk gegaan. Er werden 26 doelstellingen-uitspraken uit de totale verzameling geselecteerd. Uitgangspunt daarbij was dat een aantal doelstellingen puur inhoudelijk van karakter moest zijn en dat in een aantal een leerpsychologische component opgenomen moest zijn. Uitgaande van de oorspronkelijke verzameling doelstellingen-uitspraken werden voor twee onderdelen, nl. optellen/afrekken en vermenigvuldigen, de doelstellingen-uitspraken geselecteerd die aan deze eisen voldeden. Het resultaat was de volgende verzameling:

- 7 inhoudelijke doelen optellen/afrekken,
- 9 inhoudelijke doelen vermenigvuldigen,
- 5 doelen optellen/afrekken, met een expliciet leerpsychologische component,
- 5 doelen vermenigvuldigen, met een expliciet leerpsychologische component.⁴

De leerpsychologische component had betrekking op de vijf handelingsniveaus van Gal'perin.

Op de data van deze 26 doelstellingen-uitspraken werd een hiërarchische cluster-analyse toegepast. Voor enkele details omtrent deze analyse-techniek zie par. 6. Voor dit onderzoek betekende dit dat, op grond van de door de leerkrachten gegeven beoordeling van de doelen, de relaties tussen deze doelen zodanig werden aangegeven dat een soort boomstructuur in de doelstellingen ontstond. Eerst werden die doelstellingen-uitspraken samengenomen die in de beoordeling voortdurend aan elkaar gekoppeld bleken te zijn, d.w.z. ongeveer dezelfde beoordeling kregen door veel leerkrachten. Vervolgens werd een doelstellingen-uitspraak toegevoegd die een iets minder sterke samenhang vertoonde, etc. Het gaat hier dus niet om de absolute 'belangrijkheid' van de doelen, maar om hun onderlinge samenhang. Op deze wijze ontstonden clusters van, in de ogen van de beoordelaars i.c. de leerkrachten, bij elkaar behorende doelstellingen.

Het belangrijkste punt bij de inspectie van de resultaten betreft dan uiteraard het karakter van de op deze wijze ontstane clusters. De in de analyse

opgenomen doelstellingen zijn immers van dien aard dat de leerkracht zich bij de beoordeling van de belangrijkheid óf kan richten op de inhoudscomponent (bijv. 'vermenigvuldigen') óf ook de leerpsychologische component daarbij kan betrekken. Hierdoor zou het mogelijk zijn dat clusters zouden ontstaan die opgebouwd zijn uit puur inhoudelijke doelstellingen-uitspraken óf clusters waarin doelstellingen met een leerpsychologische component domineren.

Een dergelijk overzicht van welke soort doelstellingen volgens de leerkrachten 'bij elkaar horen' kan een indicatie vormen voor de mate waarin de leerpsychologische uitgangspunten doorwerken in het denken over het curriculum.

5. Observatie-gedeelte

5.1. Enkele inperkingen

Het observatie-gedeelte kan worden beschouwd als het belangrijkste afzonderlijke onderdeel van het onderzoek. Wanneer immers in het te evalueren curriculum gedetailleerde voorschriften worden gegeven voor het opbouwen van het onderwijsleerproces verdient een directe observatie van dit onderwijsleerproces een hoge prioriteit. Het is bekend dat veel nieuwe curricula bij hun realisatie in sterke mate afwijken van hetgeen de ontwikkelaars van dit curriculum voor ogen heeft gestaan, (Fullan 1975). In elk geval hebben de evaluatie-uitspraken over bijv. effectiviteit of bruikbaarheid van een bepaald curriculum geen enkele waarde wanneer niet eerst de mate van implementatie van dat curriculum werd vastgesteld. Het gebruiken van een observatie-instrument voor een dergelijk doel komt overeen met hetgeen Rosenshine en Furst aanduiden als 'to monitor instructional programs': 'Category systems or rating systems are used to focus upon the activities considered important for high implementation of the program', (Rosenshine en Furst 1973, p. 152). De 'activities' die door de ontwikkelingsgroep belangrijk werden geacht betroffen in de allereerste plaats het gebruik van de leerpsychologische principes die aan het curriculum ten grondslag liggen. Naast een op leerpsychologische principes gericht observatie-instrument werd een observatie-instrument ontworpen voor de meer inhoudelijke kant van het curriculum; dit laatste instrument zal hier niet verder worden besproken. De genoemde leerpsychologische principes liggen ten grondslag aan een belangrijk gedeelte van de sekwentering van het curriculum,

zowel op lesniveau als op leergangniveau. In de bij het curriculum behorende leerkracht-handleidingen wordt naar deze principes talloze malen verwezen. Wanneer de leerkracht volgens de bedoelingen van de curriculumontwikkelaars te werk gaat zal hij bij de inrichting van zijn onderwijs-leerproces van deze principes gebruik maken. Met behulp van het observatie-instrument wordt nu nagegaan in welke mate de leerkracht 'gebruik maakt van' deze principes. Voordat de operationalisering van deze principes wordt besproken kunnen over dit instrument reeds een aantal opmerkingen worden gemaakt:

(a) Met 'gebruik maken van' genoemde principes wordt hier bedoeld: het uitspreken van een of meer aaneensluitende zinnen (opmerkingen, constatering, opdrachten e.d.) gericht tot één of meer leerlingen, die kwa betekenis duidelijk verband houden met de genoemde leerpsychologische principes.

Toelichting:

- De gekozen 'coding unit' is dus niet een bepaald tijdsinterval, maar één of meer zinnen die inhoudelijk samenhangen.

- Door de aard van het verband tussen het te observeren gedrag en de registratie ervan in het observatie-systeem is een redelijke kennis van de genoemde leerpsychologische principes bij de observator onontbeerlijk. Op zichzelf is dit geen negatief te waarden eigenschap voor een observatie-instrument. Het probleem is echter of, gegeven een bepaalde basiskennis, eenduidigheid kan worden bereikt over de wijze waarop een bepaalde gedraging moet worden geregistreerd. Hier speelt het bekende probleem van de tegenstelling tussen high-inference systemen en low-inference systemen (vgl. Dunkin en Bidle 1974, p. 56). Overeenstemming werd hier bereikt door middel van het onafhankelijk van elkaar benoemen van gedragingen in video-opnamen door ontwikkelaar en evaluator, gevolgd door vergelijking en wederzijdse afstemming.

- Dat het gedrag van de leerkracht 'verband houdt met' genoemde leerpsychologische principes wordt slechts afgeleid uit de inhoudelijke betekenis van de leerkracht-uitingen. Er wordt bijv. niet nagegaan of de leerkracht de principes op theoretisch niveau 'beheerst'.

- De leerkracht kan zich richten tot de hele klas of bijv. een korte dialoog met een of enkele leerlingen aangaan. In het instrument wordt hiertussen geen verschil gemaakt.

(b) Het gaat hier om de observatie van het leerkracht-gedrag. Daarbij wordt niet nagegaan wat

de onmiddellijke invloed van dit leerkracht-gedrag is op het gedrag van de leerlingen. (Overigens wordt in tweede instantie wel degelijk de invloed van het leerkracht-gedrag nagegaan. De observatie-gegevens worden immers ook gerelateerd aan gegevens omtrent de leerling-effecten).

5.2. Operationalisering

In paragraaf 2 werden de leerpsychologische principes aangeduid die hier aan de orde zijn, nl. (1) het niveau van de handeling, (2) de mate van uitvoerigheid, (3) de graad van generalisatie en (4) de mate van beheersing. Verondersteld wordt nu dat het vrijwel onmogelijk zal zijn de tweede en de vierde parameter in observatie-studies van elkaar te onderscheiden. De 'uitvoerigheid' heeft betrekking op de mate waarin bij een bepaalde handeling alle deelhandelingen in alle volledigheid worden uitgevoerd; wanneer bepaalde deelhandelingen worden overgeslagen wordt gezegd dat 'verkorting' optreedt. De 'beheersing' slaat op de mate waarin een handeling vloeiend wordt uitgevoerd. Weliswaar is het bijv. mogelijk dat een bepaalde handeling zeer uitvoerig (dus laag op de tweede parameter) wordt uitgevoerd, terwijl toch alle deelhandelingen zeer vloeiend verlopen (dus hoog op de vierde parameter), in de praktijk zal het waarschijnlijk zeer moeilijk zijn na te gaan of een leerkracht bijv. bij een opmerking als 'probeer het nu eens sneller' aanstuurt op verkleining van de 'uitvoerigheid' of verhoging van de 'beheersing'. In het observatie-instrument worden deze parameters daarom samengenomen.

Bij de bespreking van enkele eigenschappen van de 'volwaardige mentale handeling' (Van Parreren en Carpay 1972, p. 37) wordt o.m. 'bewustheid' genoemd: 'Een derde eis is dat ze bewust worden voltrokken, waardoor ze controleerbaar blijven en de persoon er zich in principe rekenschap van kan geven'. Deze eigenschap wordt door de ontwikkelingsgroep als zeer belangrijk beschouwd. Ook op het bevorderen van deze 'bewustheid' kan het onderwijs in meerdere of mindere mate gericht zijn. Daarom is deze eigenschap als één van de relevante leerpsychologische principes in het observatie-instrument opgenomen. We hanteren dus nu de volgende categorieën: (1) niveau van de handeling, (2) bewustheid, (3) uitvoerigheid/(beheersing), (4) generalisatie.

Hieronder zal nu eerst worden aangegeven op welke wijze de genoemde vier leerpsychologische principes werden geoperationaliseerd d.m.v. het observatie-instrument.

Niveau van het handelen

Deze parameter is in het leerpsychologische model de belangrijkste en is ook in het SAC-rekencurriculum het meest uitgewerkt. Het betreft hier de bekende niveaus van handelen zoals die Gal'perins trapsgewijze vorming van begrippen aan de orde komen, (materieel, perceptief, verbaal, 'met de ogen dicht', mentaal). Het leerkrachtgedrag kan hier in drie categorieën worden ingedeeld:

- De leerkracht geeft aan op welk niveau een bepaalde handeling moet worden uitgevoerd, bijv. 'we pakken nu onze blokjes en ...' of 'kijk eens naar de getallenlijn'.
- De leerkracht probeert de handeling op een hoger niveau te laten uitvoeren, bijv. 'leg nu je blaadje op de blokjes en probeer het sommetje uit te rekenen; zeg wat je doet'.
- De leerkracht probeert de handeling op een lager niveau te laten uitvoeren. Hieronder vallen alleen die gevallen waarbij een eerdere uitvoering van de handeling op een hoger niveau niet lukte en waarbij dus tot een 'gemakkelijker' niveau wordt teruggevallen.

In bepaalde gevallen zal moeten worden uitgemaakt of een opdracht betrekking heeft op een nieuwe opgave die op een zeker niveau uitgevoerd moet worden óf op een zelfde opgave die op een hoger niveau uitgevoerd moet worden. Hier wordt slechts gesproken van 'een zelfde' opgave als de nieuwe opgave qua moeilijkheidsgraad volkomen gelijkwaardig is aan de daaraan voorafgaande.

Zoals uit het scoringsformulier (zie Figuur 2) blijkt kan hier niet alleen gelijkblijvend/hoger/lager worden aangegeven, maar ook de niveaus waarop deze manipulaties betrekking hebben. Aan de linkerzijde staat het begin-niveau, aan de bovenkant het eind-niveau. Een overgang van materieel naar perceptief wordt dus geregistreerd in de eerste rij van de tweede kolom. Leerkracht-aanwijzingen waarbij geen sprake is van overgang tussen twee niveaus komen in de diagonaal.

Mate van bewustheid

'Bewustheid' kan bevorderd worden door de leerling, wanneer hij op een bepaald niveau een handeling kan uitvoeren, nog eens nadrukkelijk erop te wijzen 'wat hij nu precies aan het doen is'. Hierdoor kan worden voorkomen dat de handeling ontaardt in een onbegrepen automatisme. Door de leerling de uitgevoerde handeling nog eens op een lager niveau te laten controleren wordt deze bewustheid bevorderd. We spreken bij dit observatie-instrument van een opdracht tot het uitvoeren van een controle-handeling als een identieke opgave op een lager

niveau moet worden uitgevoerd, terwijl in het algemeen de daaraan voorafgaande handeling op een hoger niveau eveneens correct werd uitgevoerd. Ook bij deze parameter kan in het observatie-instrument het begin- en het eind-niveau worden aangegeven. Hierbij wordt, zoals in Figuur 2 onder 'Bewustheid' te zien is, slechts de onderdriehoek gevuld: een controle-handeling wordt steeds op een lager niveau uitgevoerd dan de oorspronkelijke handeling. Een handeling die bijv. eerst op mentaal niveau werd uitgevoerd en daarna op perceptief niveau werd gecontroleerd wordt geregistreerd in de tweede kolom van de vijfde rij.

Mate van uitvoerigheid

In het algemeen zijn hier twee soorten aanwijzingen mogelijk. Enerzijds opmerkingen waarmee geprobeerd wordt de uitvoering van de handeling te detailleren of te vertragen, bijv. 'zeg nu eens precies wat je allemaal doet om ...'. Anderzijds pogingen om het leerling-gedrag te versnellen en te automatiseren. Opgemerkt kan nog worden dat deze aanwijzingen betrekking hebben op handelingen 'binnen' een bepaald niveau.

Graad van generalisatie

Deze parameter heeft betrekking op de mate waarin de handelingen van de leerlingen op specifieke objecten zijn betrokken. De leerkracht kan voorkomen dat de leerlingen de handeling te veel met specifieke objecten gaan associëren door een zelfde handeling aan gevarieerde objecten te laten uitvoeren. Als de leerkracht daarbij nog eens expliciet wijst op het identieke karakter van beide handelingen wordt dit nog eens afzonderlijk geregistreerd.

	NIVEAU					BEWUSTHEID					UITVOERIGHEID		GENERALISATIE
	ma	pe	ve	od	me	ma	pe	ve	od	me	la	an	
materieel						X	X	X	X	X			
perceptief							X	X	X	X			
verbaal								X	X	X			
ogen dicht									X	X			
mentaal										X			

Rek	
RESTCAT.	
Ander	

Figuur 2 Schema van het observatie-instrument voor de leerpsychologische component

Het scoringsformulier van het totale observatie-instrument van de leerpsychologische component ziet eruit zoals weergegeven in Figuur 2.

Wanneer zich een combinatie van een tweetal niveaus voordeed, (bijv. tegelijkertijd naar de blokjes kijkend 'met de ogen' ermee manipuleren en hardop sprekend dit manipuleren begeleiden) is steeds het laagste niveau aangehouden.

5.3. Enkele karakteristieken van het instrument

Wanneer de algemeen gangbare indeling van observatie-systemen in 'category systems' en 'sign systems' wordt gehanteerd moet dit observatie-instrument als een 'category system' worden aangemerkt: elke gedraging op een bepaalde gedragsdimensie, i.c. het verbale gedrag, wordt geregistreerd.⁵

Uit de beschrijving van de keuzen die bij het observatie-instrument gemaakt zijn blijkt duidelijk dat het niet mogelijk was gebruik te maken van reeds bestaande observatie-systemen; daarvoor zijn de te evalueren onderdelen van het curriculum te specifiek. Een daarmee samenhangend kenmerk van dit observatie-instrument is dat uitgegaan wordt van een bestaand stuk theorievorming, nl. een leerpsychologische theorie. Door een aantal auteurs wordt het vertrekken vanuit bestaande theorieën bij het ontwikkelen van observatie-systemen nadrukkelijk bepleit, (Rosenshine en Furst 1973, Nuthall 1972). Het belangrijkste voordeel lijkt daarbij te zijn dat gebruik wordt gemaakt van bestaande theoretische constructen hetgeen bevorderlijk kan zijn voor de cumulatie van kennis en verdere theorievorming. Operationalisering dienen dan wel zo veel mogelijk vanuit deze theorie plaats te vinden; waar dat niet mogelijk is dient het te worden aangegeven. De hier voorgestane werkwijze waarbij wordt vertrokken vanuit een leerpsychologische theorie is nog slechts weinig toegepast, vnl. vanuit Guilford-achtige modellen. In zijn totaliteit geeft trouwens het geheel van ontwikkelde observatie-systemen een overaccentuering van het affectieve gebied te zien ten koste van het cognitieve, (Rosenshine en Furst 1973, p. 158).

Met betrekking tot validiteitsvragen geldt dat, evenals bij het totale evaluatie-onderzoek, de inhoudsvaliditeit hier een centrale rol speelt, (vgl. Bloom 1970). Het gaat er immers niet in de eerste plaats om met behulp van de ontwikkelde instrumenten voorspellingen te doen over toekomstige criteriumgedrag of om een realisatiewijze van bepaalde theoretische constructen te vinden, maar het gaat er vooral om m.b.v. de instrumenten een 'inhoudelijk' adequaat beeld van het functioneren van het curri-

culum te krijgen. Zoals bij de meeste instrumenten waren wij ook hier voor het streven naar inhoudsvaliditeit aangewezen op een nauwkeurige analyse van het 'bedoelde' curriculum, en het in nauw overleg met de ontwikkelingsgroep samenstellen van dit instrument. Naast de inhoudsvaliditeit spelen bij een dergelijk observatie-instrument ook andere vormen van validiteit een rol. Door een aantal auteurs wordt deze validiteit vnl. gedefinieerd als de voorspellende waarde van de observatie-uitkomsten van het onderwijsleerproces voor het daaruit uiteindelijk resulterende leerling-gedrag op bijv. een eindtoets, meestal aangeduid als 'predictieve validiteit'.⁶

Om een zo groot mogelijke overeenstemming te verkrijgen tussen de wijze waarop de leerpsychologische principes in het observatie-instrument werden gehanteerd en de wijze waarop de ontwikkelingsgroep deze principes hanteerde werden aan de hand van bespreking van video-banden afspraken gemaakt over het coderen van bepaalde 'twijfelgevallen'. Daarna werd door het met tussenruimtes van drie weken observeren en coderen van een aantal video-banden getracht tot een zekere stabiliteit in de beoordeling te komen, ($r = .81$ tussen de tweede en de derde codering). In hoeverre deze stabiliteit zich in het latere onderzoek heeft gehandhaafd was niet na te gaan; het was niet mogelijk de geobserveerde lessen ook nog eens op de video op te nemen.

Een belangrijk betrouwbaarheidsprobleem bij observatie-systemen is de vraag of het de observator is gelukt een representatief beeld te krijgen van het gedrag dat hij observeren wil; dit is een betrouwbaarheidsvraag voor zover het niet-representatief zijn van het leerkrachtgedrag wordt veroorzaakt door het feit dat men in totaal te weinig tijd aan de observatie besteedt of doordat het leerkrachtgedrag instabiel is. Diverse auteurs stellen dat er geen voorschriften te geven zijn voor de tijd die men een bepaalde leerkracht moet observeren om tot een representatieve 'steekproef' uit zijn gedrag te komen. In dit onderzoek is elke leerkracht in totaal 160 minuten (4×40) geobserveerd; of dit voldoende is om toevalsfactoren uit de schakelen is niet met zekerheid te zeggen.

6. Analyse

6.1. Hiërarchische clusteranalyse

Zoals in paragraaf 4 is aangegeven is bij het doelstellingen-gedeelte van het onderzoek o.m. gebruik

gemaakt van clusteranalyse. Clusteranalyse is een met factoranalyse verwante techniek die vnl. beschrijvend van aard is. Evenals factoranalyse laat zij zich voorstellen als een manipulatie in een k-dimensionale 'testruimte', (waarbij 'k' het aantal metingen aangeeft, hier dus: het aantal leerkrachten dat de doelstellingen-uitspraken heeft beoordeeld). De objecten (hier: doelstellingen-items) kunnen worden voorgesteld als punten in deze k-dimensionale ruimte. Bij clusteranalyse probeert men 'af te lezen' uit de k-dimensionale puntenwolk welke punten (hier: doelstellingen-items) bij elkaar liggen. Een belangrijke vraag bij deze techniek is hoe 'overeenkomst' tussen twee items moet worden gedefinieerd. Hiervoor kunnen een groot aantal maten gebruikt worden, waarvan de 'euclidische afstand' het gemakkelijkst voor de stellen is. Wanneer men elk item voorstelt als een 42-vector (elke doelstelling-uitspraak wordt immers beoordeeld door 42 leerkrachten), gaat het bij deze maat om de afstand tussen twee vectorpunten. In dit onderzoek zijn we echter ook in andere wijzen van samenhang tussen de doelstellingen-uitspraken geïnteresseerd. In gevallen waarin de scores voor een tweetal items covariëren maar niet dezelfde absolute waarde hebben komen zij, wanneer de euclidische afstand als maat voor de overeenkomst wordt gebruikt, niet in hetzelfde cluster terecht. Wanneer de correlatiecoëfficiënt als maat voor de overeenkomst wordt gebruikt wordt een dergelijke samenhang uiteraard wel in rekening gebracht. Daarom is in het onderzoek, behalve de euclidische afstand, ook de correlatiecoëfficiënt als samenhangs-maat gebruikt.⁷

In paragraaf 4 is aangegeven hoe hier op een hiërarchische wijze, beginnend met de doelstellingen-uitspraken met de hoogste onderlinge samenhang, clusters van bij elkaar behorende doelstellingen worden gevormd. Dit betekent dat, naast de keuze van de wijze waarop items bij elkaar worden gevoegd, ook een keuze gemaakt moet worden voor de te volgen clusterprocedure. Bij bepaalde similarity-maten zijn slechts bepaalde clusterprocedures zinvol; een overzicht van mogelijke combinaties geeft Berkhouwer (1976), ontleend aan Wishart. Een in de literatuur vruchtbaar genoemde combinatie is die tussen de euclidische afstand als similarity-maat en de hiërarchische clusteranalyse volgens Ward. Deze methode is hier toegepast. Daarbij wordt een fusie van twee clusters tot stand gebracht op grond van de volgende procedure: binnen elk bestaand cluster wordt nagegaan wat de som van de afstanden is van de objecten uit dat cluster tot het zwaartepunt van het cluster. Clusters worden nu tot een nieuwe cluster samengevoegd op zodanige wijze dat de

toename van deze totale 'error sum of squares' minimaal is. Bij het gebruik van Pearson als similarity-maat staan een groot aantal clusterprocedures ter beschikking. De twee procedures die hier zijn gevolgd zijn de 'complete linkage' en de 'average linkage'-methode; voor een bespreking hiervan zie Berkhouwer (1976).

Er werd gewerkt met gestandaardiseerde scores. In totaal werden 26 objecten (doelstellingen-items) beoordeeld door 42 beoordelaars.

6.2. Multiple regressie-analyse

In de analyse werd o.m. per onderdeel nagegaan wat de samenhang was tussen een aantal van belang geachte variabelen en de uiteindelijke leerlingprestaties. Daarbij ging het met name om het relatieve gewicht dat aan de verschillende variabelen moest worden toegekend. In diverse bronnen wordt gewezen op multiple regressie-analyse als relevante analyse-techniek bij een dergelijke vraagstelling, (Gage en Winne 1975, p. 157; Medley, Soar en Soar 1975, p. 30). Met betrekking tot de leerpsychologische component werden een viertal variabelen als onafhankelijke variabelen gekozen⁸: (1) de mate waarin door de leerkracht van de parameter 'niveau' gebruik wordt gemaakt, (2) de mate waarin door de leerkracht van de parameter 'bewustheid' gebruik wordt gemaakt, (3) de mate waarin door de leerkracht van de parameter 'uitvoering' gebruik wordt gemaakt, (4) de mate waarin door de leerkracht van de parameter 'generalisatie' gebruik wordt gemaakt.

Deze variabelen hebben dus alle betrekking op de observatiegegevens.

7. Resultaten en conclusies

In het onderzoeksrapport worden als bijdrage tot de revisie van het curriculum een groot aantal conclusies geformuleerd en een 24-tal aanbevelingen gedaan. Datgene wat hier aan de orde zal komen heeft vnl. betrekking op het doelstellingen-gedeelte en het observatie-gedeelte van het onderzoek, met name de onderwijspsychologische component daarin.

Bij een vergelijking tussen ontwikkelaars en leerkrachten wat betreft het belang dat gemiddeld aan bepaalde groepen van doelstellingen wordt gehecht blijkt dat zich in het algemeen geen grote verschillen voordoen; zie tabel 1.

Het is daarbij opvallend dat de 'traditionele' onderdelen als optellen/afrekken en vermenigvuldigen

Tabel 1 *Gemiddelden van leerkrachten en ontwikkelaars op enkele groepen van doelstellingen*

	\bar{x} leerkr.	\bar{x} ontwikk.	t-waarde	df	p-waarde (tweez.)
optellen/aftr.	8,2	7,4	1,84	44	0,16
vermenigvuldigen	8,0	7,6	2,03	44	0,09
open beweringen	7,3	8,0	-1,69	44	0,17
materieel	8,7	8,1	1,05	44	0,37
perceptief	8,3	8,4	-0,19	44	0,86
verbaal	8,2	8,6	-1,19	44	0,30
ogen dicht	7,6	8,6	-3,45	44	0,009
mentaal	7,8	6,7	3,78	44	0,013

door de leerkrachten wat belangrijker worden gevonden, terwijl dit voor een meer specifiek SAC-onderdeel als open beweringen (evenals voor hier niet in beschouwing genomen specifieke SAC-onderdelen als kansberekening en verhoudingen) andersom is. Slechts bij twee onderdelen ontstaan significante verschillen op 5%-niveau. Het handelen 'met de ogen dicht' als een van de leerpsychologische handlingsniveaus wordt door de leerkrachten significant minder belangrijk gevonden dan door de ontwikkelaars; of dit in feite zijn oorzaak vindt in het niet voldoende begrijpen van deze stap kon uit het onderzoek niet duidelijk worden. In elk geval verdient het aanbeveling deze stap in het curriculum wat meer aandacht te geven. Dat de leerkrachten de mentale handeling belangrijker ($p = 0,013$) zouden vinden dan de ontwikkelaars was te voorzien: het traditionele rekenonderwijs verloopt immers grotendeels op dit niveau. Met de betrekking tot de 'N-doelen' kan gesteld worden dat bij alle onderdelen waarbij N-doelen geformuleerd waren, ook de leerkrachten deze doelen minder belangrijk achten dan de doelen die wél aan het SAC-rekencurriculum ten grondslag liggen. Voor de onderdelen optellen/afrekken en vermenigvuldigen⁹ gaf deze vergelijking de verschillen te zien zoals weergegeven in tabel 2.

Men zou kunnen stellen dat de leerkrachten, op enkele onderdelen na, althans op theoretisch niveau de bedoelingen van het curriculum hebben 'begrepen'. Zij kunnen bepaalde aan het curriculum ten

grondslag liggende doelen herkennen en onderschrijven deze doelen in sterker mate dan doelen waarmee de ontwikkelaars het niet eens zijn.

Bij de poging m.b.v. clusteranalyse na te gaan in welke categorieën (leerpsychologisch versus inhoudelijk) de leerkrachten voornamelijk over het rekencurriculum dachten werd gestart met clusteranalyse volgens Ward met de euclidische afstand als similarity-maat. De clusterprocedure werd uitgevoerd tot minimaal 5 en maximaal 8 clusters. In geen geval leidde deze procedure tot resultaten die in termen van de onderzoeksvraagstelling interpreteerbaar waren: alle clusterprocedures leidden tot clusters waarin voortdurend zowel leerpsychologische als inhoudelijke doelstellingen voorkwamen. Hetzelfde gold bij de clusteranalyses volgens de andere clusterprocedures. Hoewel op kleine schaal wel enkele regelmatigheden te ontdekken waren kan ter beantwoording van de onderzoeksvraagstelling de clusteranalyse-benadering als mislukt worden beschouwd. Of dit werd veroorzaakt doordat de tegenstelling tussen leerpsychologische en inhoudelijke doelstellingen hier geen vruchtbaar uitgangspunt is óf doordat de geformuleerde doelstellingen voor deze werkwijze niet adequaat zijn kon hier uiteraard niet worden vastgesteld.

In tabel 3 is weergegeven hoe vaak de verschillende leerpsychologische parameters gemiddeld per les werden gebruikt.

Het feit dat de leerpsychologische parameters bij optellen/afrekken en vermenigvuldigen nog vaker

Tabel 2 *Gemiddelden van leerkrachten op de groepen SAC-doelen en N-doelen voor de onderdelen optellen/afrekken en vermenigvuldigen*

	\bar{x} SAC-doelen	\bar{x} N-doelen	t-waarde	df	p-waarde (tweez.)
optellen/aftr.	8,2	6,6	6,16	41	< 0,001
vermenigvuldigen	8,0	6,3	4,80	41	< 0,001

Tabel 3 Aantal malen dat de diverse leerpsychologische parameters gemiddeld per les voorkwamen.

	optellen/afrekken	vermenigvuldigen	open bew.	totaal
niveau	23	20	15	58
bewustheid	10	10	7	27
uitvoerigheid	4	2	0	6
generalisatie	1	1	1	3
Totaal	38	33	23	

worden gebruikt dan bij open beweringen duidt er in elk geval op dat deze parameters zeker bij de 'traditionele' onderdelen gebruikt kunnen worden. De verschillen in totaalgebruik tussen de diverse parameters zijn niet gering, waarbij de verhouding tussen de verschillende parameters bij de drie onderdelen ongeveer gelijk is, nl. voor de parameters niveau, bewustheid, uitvoerigheid en generalisatie resp. 20 : 10 : 2 : 1.

Wanneer de ontwikkelingsgroep van mening is dat deze verdeling niet beantwoordt aan de verwachtingen die zij in dit opzicht van het gebruik van de parameters heeft zal zij in de toekomst meer nadruk op het gebruik van de twee laatstgenoemde parameters moeten leggen, bijv. door het opnemen van uitgeschreven voorbeelden van een dergelijk gebruik in het curriculum.

Wanneer deze observatiegegevens aan de hand van Figuur 2 nog wat gedetailleerder worden bekeken komen nog enkele interessante gegevens naar voren.

Het bleek namelijk dat bij de parameter 'niveau' ongeveer 1/4 van alle gebruik betrekking had op het gelijkblijvende materiële niveau en eveneens ongeveer 1/4 op het gelijkblijvende mentale niveau, (in Figuur 2 resp. de cel linksboven en cel rechtsonder van de parameter 'Niveau'). Dit betekent dat van deze parameter dus een weinig gevarieerd gebruik wordt gemaakt: veel handelingen vinden plaats óf volkomen abstract 'uit het hoofd' óf op de meest elementaire wijze, nl. alleen handelend met materiaal.

Bij de parameter 'bewustheid' bleek ongeveer 2/3 van het totale gebruik betrekking te hebben op

materiële controle, (in Figuur 2 de eerste kolom van de parameter 'Bewustheid').

Bij de regressie-analyse op de meer algemene variabelen bleek 'gebruik van leerpsychologische parameters' een positieve samenhang met leerlingenprestaties te vertonen. Zoals eerder uiteengezet werd deze variabele ook op meer gedetailleerde wijze tot voorwerp van analyse gemaakt door de vier leerpsychologische parameters als onafhankelijke variabelen in een regressie-analyse te gebruiken. Dit leverde resultaten op zoals weergegeven in Tabel 4. De samenhang is voor de diverse onderdelen niet identiek: zij is bij de onderdelen optellen/afrekken en open beweringen iets sterker dan bij vermenigvuldigen, waar op de parameter 'niveau' een zwakke negatieve samenhang te constateren is. Het is waarschijnlijk verstandig niet af te gaan op een enkele coëfficiënt.¹⁰ De algemene tendens is echter duidelijk: de samenhang met 'uitvoerigheid' en 'generalisatie' is afwezig, de samenhang met 'niveau' is zwak en niet eenduidig. Het eerder geconstateerde effect van de meer algemene variabele 'gebruik van leerpsychologische parameters' wordt vnl. veroorzaakt door de parameter 'bewustheid'. Het verdient dus waarschijnlijk aanbeveling dat de ontwikkelingsgroep in het vervolg de mogelijkheden die bij deze parameter kunnen worden benut nog meer aandacht geeft dan in het verleden is gebeurd. Met name dienen de mogelijkheden tot controlehandelingen gedetailleerd te worden aangegeven. Daarbij zal er op gewezen moeten worden dat ook op het niet-materiële niveau controlehandelingen uitgevoerd kunnen worden; deze mogelijkheid werd immers vrijwel

Tabel 4 Bèta-coëfficiënten in de regressievergelijking waarin de gemiddelde leerlingenscore is de afhankelijke variabele en de vier leerpsychologische variabelen zijn de onafhankelijke variabelen; gespecificeerd voor de verschillende onderdelen.

	niveau	bewustheid	uitvoerigheid	generalisatie
optellen/aftr.	.10	.59	-.05	-.04
vermenigvuldigen	-.28	.83	.10	-.02
open bew.	.32	.37	-.01	-.10

niet gebruikt, zoals uit de observaties bleek. Ook op grond van ander onderzoek kan overigens worden gewezen op het grote belang van deze parameter, vooral geconcretiseerd in het uitvoeren van controlehandelingen, (Van Eerde en Verhoef 1978).

In het observatie-gedeelte van het onderzoek is nagegaan in welke mate de leerkrachten in de praktijk de onderwijspsychologische component van het curriculum realiseren. Een belangrijke vraag daarbij is of dit gedrag op een bewuste en weloverwogen manier tot stand komt, m.a.w. of de leerkracht wat dit aspect van zijn handelen betreft 'zicht heeft' op zijn eigen functioneren. In een ander deel van het onderzoek is aan de hand van een aantal interviewvragen ten aanzien van de onderdelen van het curriculum nagegaan in welke mate de leerkracht bij dit onderdeel probeerde te werk te gaan volgens de aanwijzingen van het SAC-rekencurriculum. Een indicatie voor de mate waarin de leerkracht 'zicht heeft' op zijn eigen functioneren in deze kan nu wellicht worden gevonden in de samenhang tussen het eigen oordeel omtrent de mate waarin hij te werk ging 'volgens het programma' en de mate waarin dit volgens de observatiegegevens het geval was. Voor het gebruik van de leerpsychologische parameters is dit verband nagegaan (r). Bij de onderdelen optellen/afrekken, vermenigvuldigen en open beweringen was in dit verband respectievelijk 0,31; 0,32 en 0,35.

Het genoemde verband is dus slechts zeer zwak aanwezig. Dit kan waarschijnlijk niet worden verbeterd door bijv. in het curriculum nog eens nadrukkelijk op het belang van deze parameters te wijzen; op theoretisch niveau 'weet' men dit immers en, volgens de gegevens uit het doelstellingen-gedeelte, is men het daarmee volledig eens. Als meer algemene conclusie uit deze gegevens kan waarschijnlijk worden gesteld dat dit curriculum (en de gegeven begeleiding) weliswaar geschikt is om de leerkracht op theoretisch niveau vertrouwd te maken met de uitgangspunten en werkwijzen van dit rekencurriculum en vervolgens ook leidt tot een aanvaardbaar niveau van praktijktoepassing, maar dat deze toepassing gepaard gaat met een geringer inzicht in en bewustheid van eigen handelen door de leerkracht dan wenselijk is. Om hierin verbetering te brengen zouden misschien, naast een nog meer expliciet uitwerken van de psychologische principes in het curriculum, in de schoolbegeleiding gerichte trainingen moeten worden gegeven, toegespitst op bepaalde relevante gedragingen, bijv. het leren hantieren van bepaalde parameters.

De conclusies die uit het onderzoek werden getrokken en waarvan er hier enkele zijn weergegeven hadden vnl. betrekking op herziening van het curriculum. Enkele hadden echter ook duidelijk betrekking op veranderingen die de leerkracht in zijn gedrag zou kunnen aanbrengen, (uiteraard gesteund door adequate wijzigingen in het curriculum). Dit vormt een illustratie van het in de inleiding genoemde streven tot doorbreking van de scheiding die veelal nog aanwezig is tussen curriculum-onderzoek en onderzoek van het leerkrachtgedrag. Essentieel blijft hier echter steeds de nadruk die wordt gelegd op de functie die de onderwijspsychologie daarbij kan vervullen.

Noten

1. Het moeten volgen van een dergelijke procedure hangt samen met het feit dat tijdens het ontwikkelingswerk niet expliciet aandacht aan de doelformulering was geschonken. Door een aantal auteurs wordt gewezen op de taak van de evaluator dit bij de evaluatie alsnog te doen, bijv. Scriven 1974, p. 66; Popham 1975, p. 47.
2. Het erbij betrekken van 'andere' doelstellingen dient hier dus vnl. om de beoordeling van de doelstellingen van de ontwikkelingsgroep wat meer relief te geven. Deze functie is geheel afwijkend van die van de niet-cursusspecifieke doelen die in de zogenaamde Goal Free Evaluation (GFE) een rol spelen. Bij deze o.m. Cronbach voorgestane vorm van evaluatie, waarbij in de eerste plaats niet wordt nagegaan of de doelen van de ontwikkelaars bereikt worden, maar of het curriculum 'the national interest' dient (Cronbach 1963, p. 54), of waarbij wordt nagegaan wat de effecten van het curriculum zijn 'against a profile of demonstrated needs in this region of education' (Scriven 1974, p. 35), gaat men er in feite vanuit dat men de beschikking heeft over een soort totaal-overzicht van mogelijke doelen binnen een bepaald vakgebied, of nog pretentieuzer, over inzicht in een geheel van doelstellingen waarmee 'the national interest' gediend is. N.m.m. is een dergelijke opstelling gevaarlijk omdat geen enkele evaluator in werkelijkheid een dergelijke brede kijk op alle relevante doelstellingen kan hebben. En zelfs wanneer dit overzicht over alle mogelijke relevante doelen wel voorhanden zou zijn, zou het onlogisch zijn GFE aan te bevelen. Als men immers alle belangrijke doelen zou kennen, zou het meer voor de hand liggen op grond van deze doelen een curriculum te ontwikkelen en te zien of daarmee deze doelen bereikt worden dan op grond van andere, o.m. niet-belangrijke doelen een curriculum te ontwikkelen en achteraf bezielen of dit curriculum óók de belangrijke doelen bereikt. Bepaalde noties die het concept GFE heeft opgeleverd, zoals bijv. het bedacht zijn op niet-bedoelde

- neveneffecten, blijven natuurlijk bij alle evaluatie van belang.
3. Bijvoorbeeld: De beperkte reikwijdte van de op deze wijze formuleerbare doelstellingen, waardoor het aantal in feite nagestreefde doelstellingen altijd veel groter is dan het aantal dat men expliciet in zijn doelstellingen-omschrijving kan opnemen (Atkin 1968, p. 234) of zelfs de belangrijkste doelstellingen nooit in de omschrijving kunnen worden opgenomen (Trow 1970, p. 297).
 4. Voorbeeld: 'Bij het leren van tafels moet de leerling op een bepaald moment de tafelopgave kunnen uitvoeren 'met de ogen'; (3×7 is drie hoopjes van zeven met je ogen bij elkaar voegen en optellen). Deze doelstellingen-uitspraak is een operationalisatie van het construct 'perceptieve handeling' bij het onderdeel 'vermenigvuldigen'.
 5. Met de eis van Medley en Mitzel (1963, p. 299) en Dunkin en Biddle (1974, p. 73) dat met een category-system een bepaalde dimensie, een facet van het totale gedrag, in kaart moet worden gebracht, heb ik weinig moeite. De daaraan door sommigen (Herbert en Attridge 1975, p. 8) verbonden eis dat het observatie-systeem nu ook zodanige categorieën moet bevatten dat gesproken kan worden van een representatieve weergave van deze gedragsdimensie lijkt me dubieus. Het lijkt mij voldoende dat uit het instrument duidelijk wordt op welke punten geen representativiteit wordt bereikt. Dit is m.i. mogelijk door restcategorieën op te nemen; hoe meer deze restcategorieën gevuld zijn, hoe kleiner de representativiteit van het instrument ten opzichte van de gedragsdimensie. Een geringe representativiteit is daarbij zeker niet per se een negatief gegeven. Het zich concentreren op bepaalde onderdelen van bijv. de verbale gedragsdimensie maakt het juist mogelijk deze onderdelen intensief te bestuderen. Wanneer steeds geprobeerd zou worden tot een representatief beeld te komen van bijv. het 'verbale gedrag' zou men óf in oppervlakkige categorieën vervallen óf een zeer uitgebreid en onhanteerbaar observatie-systeem krijgen. Het hier besproken instrument is zeker niet als representatief voor het verbale leerkracht-gedrag te beschouwen; wel worden alle verbale leerkracht-uitingen erin opgenomen.
 6. Deze laatste aanduiding dient n.m.m. onder enig voorbehoud gehanteerd te worden. Men dient niet te vergeten dat dit begrip in de literatuur heel sterk is verbonden met het relateren van instrument (meestal test-) uitkomsten aan externe criteria, waarbij datgene wat het instrument werkelijk meet van zeer ondergeschikt belang is, zolang de predictieve waarde gegarandeerd is, (vgl. bijv. Kerlinger 1976, p. 459; De Groot 1972, p. 265). Van deze laatste opvatting kan bij formatief evaluatie-onderzoek meestal geen sprake zijn. In dit geval gaat het er wel degelijk ook om in het observatie-instrument theoretische begrippen te verwerken op een wijze die in overeenstemming is met de betekenis van deze begrippen in de oorspronkelijke theorie, althans voor zover dit mogelijk is. Dit betekent dat bij het nagaan van de relatie tussen observa-
- tiegegevens en leerlingenprestaties een bepaald aspect van de constructvaliditeit aan de orde is, niet in de betekenis van 'nagaan of het hier gebruikte construct "werkelijkheidswaarde" heeft', maar van 'nagaan of de hier gebruikte operationalisering van het construct (nl. in observatie-termen en bijbehorende co-deerregels) valide is'. (vgl. ook Popham 1975, p. 122 voor een soortgelijke redenering met betrekking tot tests). Door diverse auteurs is de nadruk gelegd op het belang van een dergelijke validering wanneer men bij zijn observatie-systeem gebruik maakt van begrippen uit bestaande, bijv. leerpsychologische, theorieën.
7. Voor het onderlinge verband tussen r^2 en euclidische afstand als similarity-maten zie Verbeek 1976, p. 240.
 8. Deze vier variabelen zijn te beschouwen als een 'uitsplitsing' van de meer algemene variabele 'gebruik van leerpsychologische parameters' in de lesituatie. Deze meer algemene variabele werd in een andere regressie-analyse als een van de onafhankelijke variabelen gebruikt, (zie Verloop 1977). Nadrukkelijk wordt er op gewezen dat de in dit artikel gerapporteerde analyse op de afzonderlijke leerpsychologische parameters tegelijk met de regressie-analyse op de meer algemene variabelen werd uitgevoerd. Om deze reden is de in een dergelijke situatie soms aanbevolen cross-validatie (McNemar 1969, p. 208 e.v.) hier niet noodzakelijk.
 9. Voor het onderdeel 'open beweringen' waren geen N-doelen geformuleerd omdat dit onderdeel in het traditionele rekenonderwijs (waaraan immers de N-doelen werden ontleend) relatief minder voorkomt.
 10. Voor het precieze effect van de afzonderlijke variabelen zou het hier uitgevoerde correlationeel onderzoek door meer experimenteel onderzoek gevolgd moeten worden, (vgl. Rosenshine en Furst 1973, p. 122). In het kader van dit evaluatie-onderzoek was dit uiteraard niet mogelijk.

Literatuur

- Anderson, R. C., R. J. Spiro en W. E. Montague (eds.), *Schooling and the acquisition of knowledge*, Hillsdale, 1977.
- Atkin, J. M., Behavioral objectives in curriculum design: a cautionary note, in: *The Science Teacher*, 1968 (35).
- Berkhouwer, Y., *Clustan IB een programma voor clusteranalyse*, Centrum voor data-analyse, Utrecht, 1976.
- Bloom, B. S., *Toward a theory of testing which includes measurement-evaluation-assessment*, in: Wittrock en Wiley, 1970.
- Cole, M. en I. Malzman (eds.), *A handbook of contemporary Soviet psychology*, New York, 1969.
- Creemers, B., *Enkele ontwikkelingen in het onderwijs-onderzoek; verslag van mijn Amerikaanse studiereis*, Inst. voor Onderwijskunde, R.U. Groningen, 1978.
- Creemers, B. en K. Meijer, *Formatieve curriculumevaluatie*, in: *Info*, 1978 (10), nr. 1, pp. 3-28.

- Cronbach, L., Course improvement through evaluation, in: *Teachers College Record*, (1963 (64), pp. 672-683. Ook in: Worthen en Sanders, 1973.
- Dunkin, M. J. en B. J. Biddle, *The study of teaching*, New York, 1974.
- Eerde, D. van en L. Verhoef, Analyse van het optellen en aftrekken op de basisschool, in: *Pedagogische Studiën*, 1978 (55), pp. 354-367.
- Edwards, K. J., Summative evaluation: some basic considerations, in: G. D. Borich (ed.), *Evaluating educational programs and products*, New Jersey, 1974.
- Eisner, E. W., *Educational connoisseurship and educational criticism: their form and functions in educational evaluation*, paper AERA-congres, 1976.
- Fullan, M., *Implementation: Its nature and determinants*, paper IMTEC-conferentie, Oslo, 1975.
- Gage, N. L. en P. H. Winne, Performance Based Teacher Education, in: K. Ryan (ed.), *Teacher Education*, NSSE, Chicago, 1975.
- Gagné, R. M. en L. J. Briggs, *Principles of instructional design*, New York, 1974.
- Gal'perin, P. J., Stages in the development of mental acts, in: M. Cole en I. Malzman (eds.), *A handbook of contemporary Soviet psychology*, New York, 1969.
- Glass, G. V., *Design of evaluation studies*, in: Worthen en Sanders, 1973.
- Groot, A. D., *Methodologie: grondslagen van onderzoek en denken in de gedragswetenschappen*, 's-Gravenhage, 1972.
- Herbert, J. en C. Attridge, A guide for developers and users of observation systems and manuals, in: *Am. Ed. Res. Journal*, 1975 (17), pp. 3-20.
- Kerlinger, F. N., *Foundations of behavioral research*, London, 1976.
- Klahr, D. (ed.), *Cognition and instruction*, Hillsdale, 1976.
- MacDonald, J. B., *The portrayal of persons as evaluation data*, paper AERA-congres, 1976.
- MacKay, D. A. en T. O. Maguire, *Evaluation of instructional programs*, Edmonton, 1971, (ERIC report ED 066 788).
- McNemar, Q., *Psychological statistics*, New York, 1969.
- Medley, D. M. en H. E. Mitzel, Measuring classroom behavior by systematic observation, in: N. L. Gage (ed.), *Handbook of research on teaching*, Chicago, 1963.
- Medley, D. M., R. S. Soar en R. Soar, *Assessment and research in teacher education: focus on Performance Based Teacher Education*, AACTE, Washington, 1975.
- Meijer, K., *Naar een meta-evaluatiemodel voor formatieve evaluatie*, paper CITO-symposium, oktober 1978.
- Nelissen, J., W. Klukhuhn en A. Vuurmans, *Informatieboek; achtergronden bij het rekenprogramma*, interne publ. SAC, Utrecht, 1978.
- Nuthall, G. A., *Ausgewählte neuere Untersuchungen zur Unterrichtsinteraktion und zum Lehrverhalten*, vert. in: Wulf, 1972.
- Parreren, C. F. van en J. A. M. Carpay, *Sovjetpsychologie aan het woord*, Groningen, 1972.
- Popham, W. J., *Educational evaluation*, Englewood Cliffs, 1975.
- Provus, M., Evaluation of ongoing programs in the public school system, in: *NSSE 68th Yearbook*, 1969 pp. 242-283.
- Rosenshine, B en N. Furst, The use of direct observation to study teaching, in: R. M. W. Travers (ed.), *Second handbook of research on teaching*, Chicago, 1973.
- Scheyer, P. T. en R. E. Stake, *A program's self evaluation portfolio*, paper AERA-congres, 1976.
- Scriven, M., The methodology of evaluation, in: *AERA Monograph Series on Curriculum Evaluation*, nr. 1, Chicago, 1967. Ook in: Worthen en Sanders, 1973.
- Scriven, M., Prose and cons about goal-free evaluation, in: W. J. Popham (ed.), *Evaluation in education*, AERA, Berkeley, 1974.
- Stufflebeam, D. L. e.a., *Educational evaluation and decisionmaking*, Bloomington, 1974.
- Trow, M., *Methodological problems in the evaluation of innovation*, in: Wittrock en Wiley, 1970.
- Verbeek, A., Clusteranalyse, in: *Mens en Maatschappij*, 1976, nr. 3, pp. 230-272.
- Verloop, N., *Formatieve curriculum-evaluatie; verslag van evaluatie-onderzoek naar het functioneren van een leergang rekenen voor klas 2 van het basisonderwijs*, vakgroep onderwijskunde, R.U. Utrecht, 1977.
- Verloop, N., *Teacher effectiveness at implementation of a curriculum based on explicit principles of learning theory*, paper 4e IAEA-conferentie, Baden, Oostenrijk, mei 1978.
- Wittrock, M. C. en D. E. Wiley (eds.), *The evaluation of instruction; issues and problems*, New York, 1970.
- Wolf, R. M., Data analysis and reporting considerations in evaluation, in W. J. Popham (ed.), *Evaluation in Education*, Berkeley, 1974.
- Worthen, B. R. en J. R. Sanders, *Educational evaluation. theory and practice*, Belmont, 1973.
- Wulf, C., *Evaluation*, München, 1972.

Curriculum vitae

N. Verloop, (geb. 1949) was na zijn onderwijzersopleiding (1970) werkzaam in het basisonderwijs. Studeerde van 1972 tot 1977 onderwijskunde aan de Rijksuniversiteit te Utrecht. Was van sept. 1977 tot maart 1979 als wetenschappelijk onderzoeker verbonden aan het Nijmeegs Instituut voor Onderwijsresearch (K.U. Nijmegen). Is sinds maart 1979 werkzaam bij de onderzoeksafdeling van het CITO. Interessen liggen op het terrein van de onderwijspsychologie, opleidingsproblematiek en leerplanevaluatie.
Adres: CITO, Oeverstraat 65, Arnhem.