

Problemen bij summatief evaluatie-onderzoek

W. VAN DE GRIFT

*Centrum voor Onderwijsonderzoek van de
Universiteit van Amsterdam*

1 Inleiding

Evaluatie-onderzoek¹ heeft zich in het recente verleden mogen verheugen in de aandacht van de Nederlandse onderwijswetenschappers. De eerste evaluatierapporten van de grote onderwijsvernieuwingprojecten die in het begin van de zeventiger jaren zijn gestart, zijn inmiddels verschenen en hebben op niet altijd gelukkige wijze de dagbladers gehaald (cf. Stokking & De Vries, 1981; Van den Berg & Welten, 1981). Ook van recenter gestarte onderwijsvernieuwingprojecten, zoals het Mavo-project, het Landelijk Leerplaatsenproject, de nieuwe Basisschool, en enkele experimenten Middenschool zijn inmiddels evaluatierapporten verschenen (cf. Creemers, 1977; Boersma & Van Viersen, 1979; Wagenaar, Van Kessel & De Vries, 1979; Van Viersen & Boersma, 1980; Bosman, Van de Grift & De Jong, 1981; Van de Grift & De Jong, 1981; De Vries, 1982; Van Eck & Lington, 1982; Coopmans, 1982; Van Veldhuizen & Span, 1982; Van de Grift, De Vries & Baltzer, 1983). Daarnaast verschenen meer beschouwende publikaties over evaluatie-onderzoek. We kunnen hier denken aan besprekingen van uitgevoerde evaluatie-onderzoekingen. De externe evaluatie van het Innovatieproject Amsterdam werd in een forum besproken en in de Informatiebladen van de Rijksuniversiteit Groningen werd hiervan verslag gedaan. Slavenburg besprak het evaluatie-onderzoek van het GEON-project (cf. Slavenburg, 1982). Behalve besprekingen van uitgevoerde evaluatie-onderzoekingen werden er ook meer algemeen beschouwende artikelen over evaluatie-onderzoek gepubliceerd. Van de kant van de Stichting voor Onderzoek van het Onderwijs (SVO) verscheen een memo over de achtergronden en uitgangspunten voor beleidsvoering bij evaluatie-

atie-onderzoek (cf. SVO, 1980). In 1981 verscheen de bundel 'Aspecten van Leerplan-evaluatie', een bundel voordrachten en papers naar aanleiding van het 10-jarig bestaan van het CITO (cf. Weeda, 1981). Het Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam onderzocht de mogelijkheden en beperkingen voor een instituutprogramma over evaluatie-onderzoek in het onderwijs (cf. Bos et al., 1982). Ook op minder institutioneel georganiseerde wijze behandelden onderzoekers onlangs evaluatiethematieken (Van de Grift, 1981; Stokking, 1982; Hofstee, 1982).

Op het eerste gezicht lijken de behandelde onderwerpen over evaluatie-onderzoek divers. De Groot (1981) bespreekt enkele zorgwekkende verschijnselen in de praktijk van evaluatie-onderzoek en stelt als remedie een probleemgerichte benadering voor. Stokking (1982) geeft een introductie in de bestaande evaluatiemodellen. Creemers (1981) laat zien dat deze nauwelijks gebruikt worden. Hofstee (1982) meent enkele pseudodilemma's bij evaluatie-onderzoek op te ruimen en stelt een nieuw evaluatiemodel voor. Stokking (1983) is vervolgens verbaasd over het door Hofstee ondernomen opruimwerk en is teleurgesteld in Hofstees alternatief. Van de Grift (1981) gaat in op de problematiek van de verschillende kennisbelangen bij evaluatie-onderzoek en Meijer (1981) gaat in op meta-evaluatie. Een aantal thematieken keren in deze publikaties regelmatig terug: de bruikbaarheid van de resultaten van evaluatie-onderzoek; het jargon van de evaluatie-onderzoekers en de kwaliteit van evaluatie-onderzoek. Er is duidelijk iets aan de hand met evaluatie-onderzoek. In dit artikel zullen we stil staan bij enkele recent onderkende problemen bij summatief evaluatie-onderzoek. We beperken ons hierbij tot evaluatie-onderzoek waarin gebruik gemaakt wordt van het experimentele model.

2 Summatief evaluatie-onderzoek: een karakteristiek

De eerste ontwikkelingen in evaluatie-onder-

zoek dateren uit de twintiger jaren, toen Amerikaanse onderzoekers geïnteresseerd raakten in het meten van leerlingprestaties. In deze periode hadden de evaluatie-onderzoekers grote aandacht voor testtheorie, testtechnologie en het vervaardigen en preciseren van meetinstrumenten voor het vaststellen van leerlingprestaties. Aan de onderwijsprogramma's zelf werd door de evaluatie-onderzoekers geen aandacht besteed (cf. Cronbach, 1963; Tyler, 1969, 399). In de periode na 1940 ontstond een grotere belangstelling voor hogere cognitieve en affectieve doelstellingen van het onderwijs (cf. Tyler, 1949). Met name het werk van Tyler staat dan centraal. Tyler vatte evaluatie op als het proces waarin bepaald wordt in welke mate de onderwijsdoelen werkelijk bereikt worden door het curriculum en instructieprogramma (cf. Tyler, 1949, 105-106). Na de lancering van de Russische sputnik (1957) ontstond in de Verenigde Staten verhoogde aandacht voor de ontwikkeling van curricula wis- en natuurkunde en voor evaluatie-onderzoek naar het effect hiervan (cf. Tyler, 1969, 1). Dit kwam ook tot uitdrukking in nieuwe bijdragen over de methodologie van evaluatie-onderzoek. We denken hier met name aan het werk van Cronbach (1963), Scriven (1967) en Bloom (1969).

In Nederland kreeg men pas veel later aandacht voor evaluatie-onderzoek. Pas in de tweede helft van de vijftiger en zestiger jaren kwam bij de psychologen een ontwikkeling van studie- en schooltoetsen op gang (cf. De Groot, 1976). De Groot definieert in deze periode evaluatie-onderzoek als: 'het methodisch en streng empirisch-wetenschappelijk, dus strict objectief nagaan, in hoeverre de methode van beïnvloeding het gewenste effect heeft (gehad).' (De Groot, 1959, 137, toevoeging in het origineel).

In deze periode raakt met name in de psychologie de term evaluatie-onderzoek ingeburgerd. Het ging hier in het algemeen om het evalueren van methoden en/of concrete pogingen om 'het gedrag van medemensen systematisch te beïnvloeden' (cf. De Groot, 1981, 20). En pas in de zeventiger jaren kreeg evaluatie-onderzoek in Nederland een steviger voet aan de grond (cf. Van Calcar, 1976; Slavenburg, 1978; Appellhof, 1979; Stokking & De Vries, 1981).

Het probleem dat ten grondslag ligt aan summatief evaluatie-onderzoek is in de woor-

den van Tyler en De Groot al aan de orde geweest. Weiss formuleerde het als volgt: 'How well is the program meeting the purposes for which it is established?' (Weiss, 1972, 2).

De klassieke aanpak van dit probleem bestaat in het beantwoorden van twee deelvragen: (in hoeverre) worden de doelen van een onderwijsvernieuingsprogramma bereikt?; en is het bereiken van deze doelen toe te schrijven aan het programma?

De eerste deelvraag wordt behandeld door het meten van de effecten, die men van het programma verwacht. Daartoe worden eerst de van het programma te verwachten effecten geoperationaliseerd. Op basis hiervan wordt een meetinstrument geconstrueerd, waarmee men deze effecten kan vaststellen. Nadat het meetinstrument tot tevredenheid onderzocht is op validiteit en betrouwbaarheid is het klaar voor afname.

De tweede deelvraag wordt aangepakt door gebruik te maken van het experimentele model. In dit model wordt gewerkt met twee groepen, die beide onderzocht worden: een experimentele groep en een controlegroep. Beide groepen worden at random (d.w.z. door het lot) samengesteld². De experimentele groep ondergaat het programma, en de controlegroep niet. Vervolgens wordt het geconstrueerde meetinstrument bij beide groepen zowel voor de start van het programma, als na afloop van het programma, afgenomen. Daarna worden de verschillen tussen beide groepen voor deze metingen vastgesteld. Het programma wordt als succesvol gezien, wanneer de experimentele groep meer betekenisvolle verbeteringen scoort ten opzichte van de beginmeting als de controlegroep.

Dit experimentele model is in wetenschappelijk opzicht een erg effectieve manier van werken. Door de grote rol die het toeval speelt in de samenstelling van beide groepen, worden verreweg de meeste mogelijkheden geëlimineerd dat andere factoren dan het programma de oorzaak zijn van het optreden van de gemeten effecten. Programma-externe factoren, zoals rijping, programma-onafhankelijke cognitieve ontwikkeling van de personen in beide groepen, maatschappelijke en lokale belemmerende en bevorderende factoren, testartefacten, meetfouten en selectieve drop-out van personen in beide groepen zijn in dit experimentele model bijna perfect onder controle (cf. Cook & Campbell, 1979; 37 ff.).

Zo rondom 1930 werd in een aantal studies die gedaan werden in de Hawthorne fabrieken van de Amerikaanse Western Electric Company een belangrijke factor ontdekt, die met dit experimentele model niet onder controle bleek te zijn. Onderzoekers ontdekten dat de aandacht en de aandacht alleen, die aan de onderzochte personen besteed werd in de experimentele conditie verantwoordelijk was voor een verhoging van de effecten. Deze effectverhoging noemen we sindsdien het 'Hawthorne-effect'³. Deze ontdekking leidde tot een bijstelling van het experimentele model. In de oorspronkelijke versie van het experimentele model onderging de controlegroep geen enkele behandeling. Om het Hawthorne-effect onder controle te houden werd aan de controlegroep een 'placebo'-programma aangeboden. Beide groepen kregen nu extra aandacht, zodat aandacht geen verklaring meer kon vormen voor de gemeten verschillen tussen de resultaten van de controlegroep en de experimentele groep.

We hebben nu een onderzoeksopzet voorliggen waarmee de klassieke dubbele vraag van summatief evaluatie-onderzoek op een bevredigende wijze beantwoord kan worden⁴. Toch vindt men deze opzet voor summatief evaluatie-onderzoek maar zelden terug in de verslagen van summatief evaluatie-onderzoek. In de meeste situaties waarin een evaluatie-onderzoeker zijn werk moet doen, doen zich namelijk nogal wat praktische problemen voor, waardoor met name het experimentele model niet volledig gerealiseerd kan worden. Bij deze praktische problemen staan we in de paragraaf kort stil.

3 *Problemen met de uitvoering van summatief evaluatie-onderzoek*

De situatie waarin een evaluatie-onderzoeker zijn werk moet doen laat zich beter karakteriseren als een turbulente actiesituatie dan als een 'clean and tidy' experiment. De veldsituatie brengt voor een evaluatie-onderzoeker talrijke problemen met zich mee, die zijn collega in het laboratorium niet heeft (cf. SVO, 1980, 25 ff.). Daarom is het evalueren van onderwijsvernieuwingsexperimenten een van de moeilijkste onderzoeksondernemingen (cf. De Groot, 1981, 12). We kunnen deze uitvoeringsproblemen bij summatief evaluatie-on-

derzoek indelen naar de soorten gevolgen die ze hebben voor het werk van de evaluatie-onderzoeker: gevolgen voor de conceptualisering van de doelen van het programma; gevolgen voor de dataverzameling en gevolgen voor de onderzoeksopzet.

De conceptualiseringsproblemen vinden hun wortels vooral in de vaak geconstateerde vaagheid van de programmadoelen (cf. Tyler, 1949, 113; Bloom, 1969, 26 ff.; Weiss, 1972). Daardoor is het vaak moeilijk om de doelstellingen te operationaliseren tot objectief constateerbare effectvariabelen. Weeda stelt in dit verband voor om de evaluator reeds in de beginfase van een leerplanontwikkelingsproject te laten participeren in de ontwikkeling. Dan kan de evaluator erop toezien dat de evaluatiecriteria operationaliseerbaar geformuleerd worden (cf. Weeda, 1981, 71). Ook bij andere auteurs zijn voor deze problematiek suggesties te vinden (cf. Bloom, 1969; Weiss, 1972).

De problemen met de toegankelijkheid van de databronnen komen veelal voort uit de bedreiging die de resultaten van evaluatie-onderzoek kunnen vormen voor het voortbestaan en de eventuele verspreiding van het geëvalueerde project. Evaluatie-onderzoek vindt immers veelal plaats in een ingewikkeld samenspel van conflictuerende belangen (cf. Van de Grift, 1981). Deze situatie brengt evaluatie-onderzoekers ertoe om zich in tal van bochten te wringen om de mogelijkheid van dataverzamelen veilig te stellen. De meest voorkomende 'bochten' zijn: evaluatie-onderzoekers proberen duidelijk te maken, dat evaluatie 'eigenlijk' ten dienste van de uitvoerenden zou moeten staan; extra aandacht voor het proces dat de uitvoerenden doormaken; en een minder 'hard nosed' presentatie van de onderzoeksplannen.

De problemen met de onderzoeksopzet komen voort uit de geringe invloed die evaluatie-onderzoekers hebben op de samenstelling van de groep die het programma ondergaat: de afwezigheid van faciliteiten of mogelijkheden om met controlegroepen te werken; en het te laat ingeschakeld worden van evaluatie-onderzoekers, waardoor voormetingen niet meer mogelijk zijn. Met name op deze groep van problemen voor de onderzoeksopzet is door evaluatie-onderzoekers in het verleden met creativiteit en inventiviteit gereageerd. Dit is tot uitdrukking gekomen in een aantal variaties op het experimentele model, waarmee en de

effecten van een programma vastgesteld konden worden en met wisselend succes een groot aantal programma-externe factoren onder controle gehouden konden worden. Campbell en Stanley analyseerden de eigenschappen van een aantal experimentele opzetten, en vervaardigden bovendien enkele nieuwe opzetten die geschikt leken voor evaluatie-onderzoek (cf. Campbell & Stanley, 1963; Cook & Campbell, 1979).

Deze drie groepen uitvoeringsproblemen hebben in voorkomende gevallen gedwongen tot afwijkingen van de klassieke aanpak van summatief evaluatie-onderzoek. Toch vormen deze uitvoeringsproblemen op zich geen voldoende reden om het streven naar de evaluatie-aanpak zoals Tyler, De Groot en Weiss die voorstaan op te geven. De drie groepen uitvoeringsproblemen hebben geleid tot nieuwe conceptualiseringsaanpakken, tot nieuwe strategieën voor het toegankelijk maken van data en vooral tot interessante en bruikbare varianten van het experimentele model. Deze drie groepen oplossingen hebben de 'gereedschapskist' van de summatieve evaluatie-onderzoeker verrijkt zonder dat de essentie van de aanpak van summatief evaluatie-onderzoek aangetast hoefde te worden. Man kan immers van deze oplossingen gebruik maken terwijl men summatief evaluatie-onderzoek toch blijft definiëren als een reeks effectmetingen, waarbij men andere verklaringen voor de opgetreden effecten, dan het vernieuwingsprogramma, probeert te elimineren door gebruik te maken van (een variant van) het experimentele model.

Toch zijn wij van mening dat de klassieke aanpak van summatief evaluatie-onderzoek aanvulling behoeft. Daarop gaan we nu in.

4 *Nieuwe problemen bij summatief evaluatie-onderzoek*

De laatste jaren worden er meer en meer problemen gesignaleerd, die de krachten van de klassieke aanpak van summatief evaluatie-onderzoek te boven gaan. Deze problemen komen voort uit de heel opmerkelijke en steeds vaker, internationaal, voorkomende constatering dat veranderingspogingen niet leiden tot verbeterde leerlingresultaten (cf. Mort, 1964, 318; Rogers & Shoemaker, 1971; Fullan & Pomfret, 1977, 337; Berman & McLaughlin, 1978, 4; Bauer & Rolff, 1978, 219; Welch,

1979.) In Nederland werd een vergelijkbare constatering gemaakt door Van Viersen en Boersma. Hun eindconclusie luidde: 'dat op de door ons onderzochte aspecten vrijwel geen verschillen voorkomen tussen leerlingen van experimenteerscholen en leerlingen van mavo-scholen buiten het Mavo-project'. (Van Viersen & Boersma, 1980, 47).

Met summatief evaluatie-onderzoek zoals we dat in het voorafgaande gekarakteriseerd hebben krijgt men nauwelijks inzicht in dit probleem van de geringe doeleffectiviteit van programma's. Dat kan ook niet. Met de klassieke vorm van summatief evaluatie-onderzoek kan men alleen maar, en dan nog in het meest ideale geval, het vernieuwingsprogramma aanwijzen als verklaring voor het wel of niet optreden van de beoogde (leerling)effecten. Wanneer de evaluatie-onderzoekers wel systematisch positieve leerlingeffecten hadden vastgesteld, dan was er niets aan de hand geweest. Maar juist omdat er zo vaak teleurstellende leerlingresultaten vastgesteld moesten worden, was het niet meer voldoende om dit af te doen met 'het onderwijsvernieuingsprogramma werkt niet'. Er ontstonden verderliggende vragen: 'Waarom werkt het onderwijsvernieuingsprogramma niet?', 'Deugt het vernieuwingsprogramma niet?', of: 'Is het vernieuwingsprogramma wel in de praktijk gerealiseerd?'. 'En waar kwam dit dan door?' 'Heeft de nameting van de effecten misschien te vroeg plaatsgevonden?' Om deze verder liggende vragen te kunnen beantwoorden ontstonden er twee nieuwe elkaar aanvullende lijnen in het denken over de opzet van summatief evaluatie-onderzoek. Eén lijn van denken kwam voort uit de zorg voor de kwaliteit van de innovatie-inhouden en de andere lijn van denken werd geïnspireerd door de vraag of de onderwijsvernieuingsprogramma's wel in de praktijk gerealiseerd waren. We zullen deze beide lijnen van denken in de volgende paragrafen behandelen.

5 *Het vergroten van de kwaliteit van de vernieuwingsprogramma's*

Zeker aanvankelijk weten de evaluatie-onderzoekers de teleurstellende leerlingresultaten, die ze vastgesteld hadden, vooral aan de kwaliteit van de vernieuwingsprogramma's. Met name onder invloed van een klassiek geworden

artikel van Cronbach kregen evaluatie-onderzoekers aandacht voor het verbeteren van vernieuwingsprogramma's (cf. Cronbach, 1963). Cronbach ging er vanuit, dat prestatie metingen aan het einde van een vernieuwingsproject weinig konden bijdragen aan de verbetering van een programma. Daarom zocht hij naar mogelijkheden om evaluatie te integreren in het proces van de ontwikkeling van vernieuwingsprogramma's. De rol van evaluatie-onderzoek zou volgens hem minder op effectmeting aan het eind, en meer op het verbeteren van vernieuwingsprogramma's gericht moeten zijn.

5.1. *Formatief evaluatie-onderzoek*

Cronbachs idee werd in 1967 door Scriven verder uitgewerkt. Hij formuleerde een nieuwe rol voor evaluatie-onderzoek: formatief evaluatie-onderzoek. Formatief evaluatie-onderzoek kan men het beste zien als effect-onderzoek in intermediaire fasen van de ontwikkeling van een vernieuwingsprogramma (cf. Scriven, 1967, 51). Het kenmerkende van formatief evaluatie-onderzoek is, in tegenstelling tot wat vaak gedacht wordt, niet dat het aan lagere methodologische eisen moet voldoen dan summatief evaluatie-onderzoek. Het kenmerkende van formatief evaluatie-onderzoek is op de eerste plaats dat het een andere rol vervult dan summatief evaluatie-onderzoek. De rol van formatief evaluatie-onderzoek bestaat in het ontdekken van de sterke en zwakke kanten van intermediaire versies van een nieuw curriculum (cf. Scriven, 1967, 51). De werkwijze bij formatief evaluatie-onderzoek is in methodologisch opzicht gelijk aan die in summatief evaluatie-onderzoek (cf. Meijer, 1981). Alleen gaat het nu niet om een onderzoek naar de definitieve versie van een vernieuwingsprogramma, maar om een onderzoek naar een conceptversie. Met het formatieve evaluatie-onderzoek, zoals Scriven dat formuleerde is er dus in vergelijking met summatief evaluatie-onderzoek weinig nieuws onder de zon. Nog steeds blijft het oude argument van Cronbach (1963) van kracht, dat een effecttoets alléén niet bepaald de geëigende manier is om programma's te verbeteren. Op zijn best leidt een observatie van teleurstellende effecten tot de conclusie dat het anders moet? De vraag bij verbeteren is echter: *hoe* dan wel anders?

5.2. *Analytisch evaluatie-onderzoek*

De klassieke aanpak van summatief evaluatie-onderzoek is steeds gericht geweest op het uiteindelijk empirische succescriterium: de mate waarin een vernieuwingsprogramma er in slaagt om de doelen te bereiken waarvoor het ontworpen is. In het voorafgaande hebben we gezien dat van veel vernieuwingsprogramma's vastgesteld moest worden dat de doeleffectiviteit ervan gering is. Dit maakt het belangrijk om vernieuwingsprogramma's op hun kwaliteit te beoordelen, nog voordat er leerlingen aan 'gewaagd' worden. Het wordt nog maar zelden gedaan, maar het is heel goed mogelijk om belangrijke kwaliteitsaspecten van een vernieuwingsprogramma te beoordelen nog voordat dat programma in de werkelijkheid wordt uitgetoetst. We spreken dan van analytisch evaluatie-onderzoek. Zo'n analytisch evaluatie-onderzoek is uitgevoerd in een deelstudie van het onderzoek ter externe evaluatie van het Landelijk Leerplaatsenproject. We zullen nu heel kort de criteria beschrijven die we in dit analytisch evaluatie-onderzoek hanteerden. De geïnteresseerde lezer vindt een meer gedetailleerde uiteenzetting en toepassing van deze criteria in ons onderzoeksrapport (cf. Van de Grift & De Jong, 1981).

Het belangrijkste criterium dat we hanteerden is dat het gebruik van een vernieuwingsprogramma gerechtvaardigd moet kunnen worden met een werkzaam principe, waarin een verband gelegd wordt tussen het vernieuwingsprogramma en de doelen ervan. Een vernieuwingsprogramma definiëren we als datgene, waarmee men de discrepantie tussen een geobserveerde en een gewenste situatie beoogt te verkleinen. Een vernieuwingsprogramma moet dan ook niet verward worden met de vernieuwingsdoelen. De doelen zijn niet meer, maar ook niet minder dan een formulering van de gewenste situatie. Een vernieuwingsprogramma geeft een antwoord op een hoe-vraag. Het is het geheel van activiteiten waarmee we vernieuwingsdoelen denken te bereiken en niet de doelen zelf. Een vernieuwingsprogramma is succesvoller naarmate het de kloof tussen de geobserveerde en de gewenste situatie meer verkleint. Om aan dit succescriterium te kunnen voldoen moet een vernieuwingsprogramma onder andere op een werkzaam principe berusten, waarin een verband gelegd wordt tussen de activiteiten, materialen en dergelijke van het programma en de doelen ervan.

Het vernieuwingsprogramma moet een bevredigende verklaring kunnen vormen voor het optreden van de gewenste situatie. Wanneer er geen theoretisch verband bestaat tussen het programma en de doelen ervan, dan is er geen reden om te verwachten dat die vernieuwingsdoelen met dat programma bereikt worden. En dan is er ook geen reden om dat programma in praktijk te brengen. We weten dan immers niet waar dat vernieuwingsprogramma toe leidt. Hiervan uitgaande zijn de eerste vragen die we aan een vernieuwingsprogramma kunnen stellen, nog voordat tot uitvoering wordt overgegaan: 'Wat is het werkzame principe van het programma?' en 'Legt dit principe een plausibel theoretisch verband tussen de activiteiten en de doelen van het vernieuwingsprogramma?' Wanneer deze vragen bevredigend beantwoord zijn kunnen we ons vervolgens afvragen: Zijn er (reeds) onderzoeksverslagen beschikbaar waaruit blijkt dat het principe inderdaad werkzaam is? Immers bij de aanwezigheid van valide contra-evidentie zou het onverstandig zijn om tot uitvoering of verspreiding van het betreffende vernieuwingsprogramma over te gaan.

Het zojuist behandelde primaire kwaliteitscriterium is echter niet het enige analytische criterium waaraan een programma moet voldoen voordat het verstandig lijkt om het te gaan uitproberen. Een programma moet ook nog uitvoerbaar zijn voor leerkrachten in schoolorganisaties. Daartoe moet het programma aansluiten bij de kennis, vaardigheden, en waarde-oriëntatie van de leerkrachten en de mogelijkheden, de organisatiestructuur en het machtsbereik van de scholen, die ermee aan de slag gaan. Uiteraard is de uitvoerbaarheid van een programma afhankelijk van de specifieke kenmerken van de veldsituatie en de projectstaf. Strikt genomen zou dit dan ook steeds ter plaatse onderzocht moeten worden. Toch valt hier in algemene zin reeds heel wat over te zeggen. Er is de laatste jaren nogal wat onderzoek gedaan naar de voorwaarden waaraan een onderwijsvernieuwingsprogramma moet voldoen voor een succesvolle uitvoering. Aan deze onderzoeksresultaten kunnen we criteria ontlenen waaraan nieuwe onderwijsprogramma's zouden moeten voldoen om het keurmerk 'uitvoerbaar' te krijgen. Het gaat hier om twee reeksen criteria, de adopteerbaarheid en de implementeerbaarheid.

De adopteerbaarheidscriteria betreffen een

aantal kenmerken van een vernieuwingsprogramma, die de beslissing van leerkrachten en scholen positief beïnvloeden om met een vernieuwingsprogramma aan de slag te gaan. Deze thematiek heeft in een groot aantal onderzoeken centraal gestaan. Deze onderzoeken zijn ondermeer samengevat in de overzichtswerken van Rogers en Shoemaker (1971) en Zaltman en Duncan (1977). Op grond hiervan kunnen we de volgende criteria opstellen waaraan een vernieuwingsprogramma moet voldoen om het keurmerk 'voor leerkrachten adopteerbaar' te krijgen: Rogers en Shoemaker noemen: een verandering en de resultaten ervan moeten 'zichtbaar' zijn; een verandering moet op beperkte schaal *uitgeprobeerd* kunnen worden; een verandering moet *verenigbaar* zijn met bestaande waarden; en een verandering moet *relatief voordeel* hebben boven het bestaande (cf. Rogers & Shoemaker, 1971, 136-157). Zaltman en Duncan noemen behalve de bovengenoemde factoren ook nog de volgende: de *impact* die een verandering heeft op sociale relaties, het gemak waarmee over een innovatie-inhoud *gecommuniceerd* kan worden; de *tijd* die het kost om een innovatie te introduceren; het risico en de *onzekerheid* die een innovatie oproepen; de *betrokkenheid* die een verandering vereist; en de ontvankelijkheid voor *successieve verandering* die voor specifieke innovaties nodig is (cf. Zaltman & Duncan, 1977, 13 ff.). Verder hebben Doyle en Ponder er op gewezen, dat leerkrachten bij hun beoordeling van vernieuwingen vooral letten op *praktische bruikbaarheid*. Het zou dan gaan om het instrumentele karakter van de vernieuwing; en de mate waarin een veranderingsvoorstel overeenkomt met de eigen situatie, en om de verhouding tussen de hoeveelheid inspanning om de verandering in te voeren en de verwachte opbrengst ervan. (cf. Doyle & Ponder, 1978.)

De implementeerbaarheidscriteria betreffen een aantal kenmerken van een vernieuwingsprogramma die het in praktijk brengen ervan positief beïnvloeden. Uit een literatuuronderzoek van Janssens is gebleken, dat niet alle hierboven genoemde kenmerken van een vernieuwingsprogramma ook voor de implementatiefase relevant zijn. Volgens Janssens zijn de volgende van de hierbovengenoemde kenmerken ook voor de implementatiefase van belang: het *relatieve voordeel* dat een verandering boven het bestaande heeft; de *verenigbaarheid*

van de innovatie met bestaande waarden; de *complexiteit* van de innovatie en de mate waarin de leerkracht een *duidelijk beeld* heeft van de innovatie (cf. Janssens, 1980, 235). Deze beide laatste factoren worden ook door Fullan en Pomfret in hun overzichtsartikel benadrukt. Volgens hen dient een vernieuwingsprogramma voor een goed verlopend implementatieproces op de eerste plaats *helder* en *expliciet* geformuleerd te zijn. Fullan en Pomfret parafraserend, zouden we de volgende hypothesen kunnen opstellen: Als een vernieuwingsprogramma weinig expliciet en weinig helder is, dan heeft dat verwarring en frustratie van de gebruiker tot gevolg. Dit leidt weer tot een lage mate van implementatie. Een tweede karakteristiek, die vitaal lijkt voor het implementatieproces, is de mate van *complexiteit*, of de moeilijkheid in het gebruik. Fullan en Pomfret formuleren hierover de volgende hypothese: 'The more difficult the change, or the greater the degree of new learning entailed by it, the more likely that the degree of implementation will vary across groups of users.' (Fullan & Pomfret, 1977, 370.)

Uiteraard beïnvloedt de mate van complexiteit de explicietheid. Immers, hoe complexer een vernieuwingsprogramma is, hoe moeilijker het is om expliciet te zijn over de operationele karakteristieken van dat programma.

6 Implementatie-onderzoek

In 1962 wezen Hyman, Wright en Hopkins er op dat 'The answer to why a program was ineffective may even reduce to the simple fact that it was not in reality operative: it existed only on paper.' (Hyman, Wright and Hopkins, 1962, 74.) In de jaren daarna kregen evaluatieonderzoekers expliciet aandacht voor het vraagstuk of programma's wel in de praktijk gerealiseerd waren. In 1967 maakte Scriven een onderscheid in drie soorten processtudies: studies waarin datgene wat er werkelijk in de klaspraktijk gebeurt omwille van de praktijk zelf bestudeerd werd; studies waarin de causale claims van het proces onderzocht werden; en formatieve evaluatie-studies, waarin vroege varianten of prototypes van nieuwe curricula werden uitgetest (cf. Scriven, 1967, 49 ff.). De tweede door Scriven onderscheiden variant is van belang voor ons probleem in deze para-

graaf. In dit type processtudie stelt een evaluatie-onderzoeker vast in welke mate een vernieuwingsprogramma gerealiseerd is. Gelijktijdig met deze klassiek geworden studie van Scriven, publiceerden Hyman en Wright een artikel waarin zij nogmaals waarschuwden voor het evalueren van papieren intenties (cf. Hyman & Wright, 1967, 744 ff.).

Aldus ontstond er aandacht voor het ontwikkelen van evaluatiemodellen, die procedures bevatten voor het beschrijven van innovatieprocessen. Hierbij kunnen we denken aan verschillende bijdragen van Stake (1967); aan het context-, input-, proces- en produktmodel van Stufflebeam (1969); en aan het discrepantiemodel van Provus (1971). Uit de aandacht die er in deze modellen besteed werd aan het innovatieproces, mogen we echter niet afleiden, dat evaluatie-onderzoekers voldoende aandacht besteedden aan onderzoek naar de implementatie van de vernieuwingsprogramma's. Nog in 1973 moesten Charters en Jones het risico benadrukken van het waardeeren van 'non-events' bij programma-evaluatie. Ook in een recent uitgevoerd Nederlands evaluatie-onderzoek lijkt het erop dat papieren intenties geëvalueerd zijn. Zo wordt door Westerlaak in het voorwoord van het al eerder genoemde onderzoeksrapport van Van Vierssen en Boersma opgemerkt: 'De lezer moet bij het beoordelen van de onderzoeksresultaten steeds in het oog houden dat het hier gaat om een momentopname in de *beginfase* van het Mavo-project (...). In dit onderzoek wordt geen eendoordeel over het Mavo-project uitgesproken, daarvoor is het nog te vroeg.' (Van Vierssen & Boersma, 1980, II.)

Om een lang verhaal kort te maken. Het heeft weinig zin om een nameting voor de doelen van een vernieuwingsprogramma te houden, wanneer men niet zeker weet, dat het vernieuwingsprogramma geïmplementeerd is. Immers wanneer de onafhankelijke variabelen niet in werking zijn dan heeft men weinig reden om te verwachten dat de afhankelijke variabelen zullen optreden.

Om de mate van implementatie van een vernieuwingsprogramma vast te stellen, zal een evaluatie-onderzoeker, net als bij effectonderzoek een meetinstrument moeten ontwerpen en ijken. Hoe dit meetinstrument eruit moet zien, wordt bepaald door het vernieuwingsprogramma. We zullen deze problematiek verhelderen aan de hand van een betrekkelijk een-

voudig voorbeeld. We gaan uit van het volgende principe: 'Wanneer ouders participeren in het opstellen van het schoolwerkplan van een school, dan vindt er een betere aansluiting plaats tussen de pedagogische doelen van de ouders en de school'. Het gedeelte na de komma is het doel van het vernieuwingsprogramma. Het bepalen van de effecten moet uiteraard ook gebeuren. Dit gaat analoog met het behandelde in paragraaf twee. Het eerste gedeelte van het juist genoemde principe is het te implementeren gedeelte, het vernieuwingsprogramma. Het vaststellen van de mate van implementatie van het vernieuwingsprogramma is in dit geval het beantwoorden van de vraag 'Participeren de ouders in het opstellen van het schoolwerkplan?' Om deze vraag goed te kunnen beantwoorden zal een evaluatie-onderzoeker bijvoorbeeld observaties op school moeten gaan maken. We zullen enkele voorbeelden geven van de aantekeningen die een evaluatie-onderzoeker tijdens zijn observaties gemaakt zou kunnen hebben.

- A. 'Ik heb een brief gevonden waarin de ouders uitgenodigd worden om deel te nemen aan het overleg op school over wel of niet gebruiken van Wiskobas'.
- B. 'Ik was bij een overleg op school, waarin leerkrachten en ouders overlegden over een verandering in het rekenleerplan'.
- C. 'Opvallend was, dat het schoolhoofd zich tijdens de discussies niets aantrok van de kritiek van de ouders op het schoolbeleid'.
- D. 'Later bleek dat de kritiek van de ouders op het beleid van de school door de leerkrachten en het schoolhoofd doorgenomen werd'.

Al deze observatie-uitspraken laten zien, dat er op deze school gewerkt wordt aan het realiseren van het vernieuwingsprogramma. Het wordt gecompliceerder als we observatie-uitspraak C. nog eens nader bezien. Hier blijkt dat op deze school het vernieuwingsprogramma optimaler gerealiseerd had kunnen worden. Welke mate van implementatie is eigenlijk optimaal en hoe stellen we dat vast? Om deze vragen te kunnen beantwoorden moeten we het vernieuwingsprogramma gaan uitschrijven in termen van de gedragingen die we kunnen verwachten als leerkrachten en ouders met dit vernieuwingsprogramma aan de slag gaan. Op basis hiervan kunnen we dan een checklist construeren waarmee we kunnen vaststellen welke van de verwachte gedragingen ook wer-

kelijk voorkomen. We kunnen nu zowel iets zeggen over de kwantiteit als over de kwaliteit van de implementatie van het vernieuwingsprogramma. De kwaliteit van implementatie is de hoeveelheid van deze gedragingen die we hebben kunnen constateren en de kwantiteit van implementatie is de frequentie waarmee we elk van deze gedragingen hebben kunnen constateren (cf. Gross, Giacuinta & Bernstein, 1971, 16). Nu kunnen we op twee manieren iets zeggen over de mate van implementatie van het vernieuwingsprogramma. Een procedure die hier erg veel op lijkt is de procedure voor het identificeren van innovatieconfiguraties, die ontworpen is door Hall en Loucks (cf. Hall & Loucks, 1978; Heck, Stiegelbauer, Hall & Loucks, 1981). Deze procedure is in Nederland en België geïntroduceerd door Van der Perre (1981). We zouden nog verfijnder kunnen observeren als we ook nog kijken naar de manier waarop het vernieuwingsprogramma gebruikt wordt. Gaat het heel mechanisch, gaat het routinematig, of hebben de leerkrachten hun omgaan met ouders al helemaal geïntegreerd in hun dagelijks bezig zijn. Ook hiervoor zijn meetinstrumenten ontworpen (cf. Hall & Loucks, 1975, 1977; Van den Berg & Vandenberghe, 1981). Het zou echter te ver voeren om alle meetinstrumenten om de mate van implementatie te bepalen, waarover evaluatie-onderzoekers inmiddels beschikken, hier te beschrijven. Een goed overzicht hiervan is te vinden in het overzichtsartikel van Fullan en Pomfret (1977) en in het artikel van Revicki, Rubin en Stuck uit 1981.

7 *Formatief evaluatie-onderzoek revisited*

In het voorafgaande hebben we gesteld dat formatief evaluatieonderzoek, op de wijze zoals Scriven dat beschreef, eigenlijk geen aanwijzingen geeft over hoe een vernieuwingsprogramma te verbeteren valt. Wanneer we formatief evaluatie-onderzoek aanvullen met implementatie-onderzoek dan kunnen we deze problematiek opnieuw bezien. Het is immers dan heel goed mogelijk om analyses uit te voeren op de relaties tussen de gevonden afzonderlijke effecten en de afzonderlijke gerealiseerde componenten van een vernieuwingsprogramma. Hiervoor is de methode van covariantiestructuuranalyse, LISREL, geschikt (cf.

Saris, z.j.). Hiermee kunnen we de correlaties bepalen tussen de afzonderlijke effecten en de afzonderlijke componenten van een gerealiseerd vernieuwingsprogramma. Op grond hiervan kunnen we de zwakke componenten van een vernieuwingsprogramma elimineren of minder aandacht geven en de sterke componenten extra benadrukken of meer ruimte geven in het vernieuwingsprogramma.

8 Conclusie

De vraag die ten grondslag ligt aan summatief evaluatie-onderzoek luidt als volgt: hoe goed slaagt een onderwijsvernieuwingsprogramma erin om de doelen te bereiken waarvoor het ontworpen is? De klassieke aanpak van dit probleem bestaat uit het behandelen van twee deelvragen: – (in hoeverre) worden de doelen van een onderwijsvernieuwingsprogramma bereikt?; – en is het bereiken van deze doelen toe te schrijven aan het onderwijsvernieuwingsprogramma? We hebben laten zien, dat het beantwoorden van alleen deze twee deelvragen niet langer als bevredigend beschouwd kan worden om de centrale vraag van summatief evaluatie-onderzoek te behandelen.

Het uiteindelijke succescriterium voor een onderwijsvernieuwingsprogramma is de mate waarin het in de praktijk gebrachte vernieuwingsprogramma er in slaagt om de kloof tussen de gewenste eindsituatie en de geobserveerde beginsituatie te verkleinen. Om aan dit uiteindelijke succescriterium te voldoen moet minimaal aan alle hierondergenoemde vereisten *tesamen* en *zonder uitzondering* worden voldaan:

1. het onderwijsvernieuwingsprogramma moet in een plausibel theoretisch verband staan met de beoogde vernieuwingsdoelen;
2. het onderwijsvernieuwingsprogramma moet werkelijk in praktijk gebracht zijn;
3. de vernieuwingsdoelen moeten werkelijk zijn opgetreden;
4. andere verklaringen voor het optreden van de vernieuwingsdoelen, dan het onderwijsvernieuwingsprogramma, moeten uitgesloten zijn.

Voor de praktijk van evaluatie-onderzoekers betekent dit dat de klassieke constituerende deelvragen van summatief evaluatie-onderzoek (vereiste drie en vier) aangevuld moeten worden met analytisch evaluatie-onderzoek

(vereiste één) waarin onderzocht wordt of er een plausibel theoretisch verband bestaat tussen het vernieuwingsprogramma en de doelen ervan en implementatie-onderzoek (vereiste twee) waarin onderzocht wordt of en in hoeverre het vernieuwingsprogramma in praktijk gebracht is. Tevens verdient het aanbeveling om nog voordat een onderwijsvernieuwingsprogramma gerealiseerd gaat worden te onderzoeken of het betreffende programma wel uitvoerbaar is voor leerkrachten in hun schoolorganisatie.

Noten

1. Met dank aan P. Appelhof, P van den Broek, M. van de Vall en W. Wardekker voor hun commentaar op een eerdere versie van dit artikel.
2. Uiteraard is dit gelijkhouden van beide groepen door middel van kans aan assumpties gebonden. Het heeft niet veel zin om hier op in te gaan, omdat een sterker alternatief niet geboden kan worden.
3. Cook rapporteerde in 1968 op basis van een grondig literatuuronderzoek en een empirische studie enige twijfels over het bestaan van het Hawthorne-effect. Dit onderzoek wordt door evaluatie-onderzoekers, die dit effect om wat voor reden dan ook niet onder controle konden houden, graag geciteerd om het Hawthorne-effect te bagatelliseren.
4. Let wel, het zou voorbarig zijn om te beweren, dat in het experimentele model alle storende factoren onder controle zijn. Zo blijft bijvoorbeeld het leereffect dat door testen optreedt, een lastig probleem (cf. Sandbergen, 1968).

Literatuur

- Appelhof, P. N., *Begeleide onderwijsvernieuwing-evaluatie van een curriculum-innovatie gericht op differentiatie van het aanvankelijk leesonderwijs*. Tilburg: Zwijzen, 1979.
- Bauer, K. O. & H. G. Rolff, *Vorarbeiten zu einer Theorie der Schulentwicklung*. In: K. O. Bauer & H. G. Rolff, *Innovation und Schulentwicklung*. Weinheim: Beltz, 1978.
- Berg, A. van den & V. J. Welten, *Externe evaluatie van het Innovatieproject Amsterdam*, Het innovatieproject gewogen, samenvattend eindverslag, conclusies en aanbevelingen van een extern evaluatie-onderzoek. Nijmegen: Hoogveld Instituut, Amsterdam: Kohnstamm Instituut, 1981.
- Berg, R. van den & R. Vandenbergh, *Onderwijsinnovatie in verschuivend perspectief*. Tilburg: Zwijzen, 1981.

- Berman, P. & M. W. McLaughlin, *Federal programs supporting educational change*, vol. VIII: Implementing and sustaining innovations. Santa Monica: Rand, 1978.
- Boersma, P. F. A. & D. J. van Viersen, *Met het Mavoproject onderweg 1*, een onderzoek onder mavo-docenten van scholen binnen en buiten het Mavoproject. Nijmegen: ITS, 1979.
- Bos, D. J., W. van de Grift, F. Riemersma & E. Roede, Bruikbaar evaluatie-onderzoek, mogelijkheden en beperkingen voor een programma over evaluatie-onderzoek in het onderwijs. *SCO-cahier*, nr. 5., 1982.
- Bosman, P., W. van de Grift & M. de Jong, *De externe evaluatie van het Landelijk Leerplaatsenproject*, deel I: Beleidscontext en voorgeschiedenis. Amsterdam: Kohnstamm Instituut van de Universiteit van Amsterdam, 1981.
- Bloom, B. S., Some theoretical issues relating to educational evaluation. In: R. W. Tyler, *Educational evaluation: new roles new means* (NSSE). Chicago: The university of Chicago press, 1969.
- Calcar, C. van, *Innovatieproject Amsterdam*, deel II: Tussenstand. Amsterdam: Van Gennep, 1976.
- Campbell, D. T. & J. C. Stanley, *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1963.
- Charters, W. W. jr. & J. E. Jones, On the risk of appraising non events in program evaluation. *Educational researcher*, 1973, vol. 2, no. 11.
- Cook, T. D. & D. T. Campbell, *Quasi-experimentation, design & analysis. Issues for fieldsettings*. Chicago: Rand McNally, 1979.
- Coopmans, J., *De middenschool uit*, een tussentijds verslag van een onderzoek bij leerlingen die de middenschool afsluiten, band 1, Verantwoording en onderzoeksresultaten. Nijmegen: ITS, 1982.
- Creemers, H. P. M. (red.), *Beschrijving beginsituatie experimenten middenschool I*. Haren: Rion, 1977.
- Creemers, B., *Curriculumevaluatie: betere theorie, betere praktijk*. In: P. Weeda, *Aspecten van leerplanevaluatie*. 's-Hertogenbosch: Malmberg, 1981.
- Cronbach, L. J., Course improvement through evaluation. *Teachers college record*, 1963, vol. 64, no. 8.
- Doyle, W. & C. A. Ponder, The practicality ethic in teacher decision making. *Interchange*, 1978, vol 8, nr. 3.
- Eck, E. van & H. Lington, *Middenschool in de maak*. Amsterdam: Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam, 1982.
- Fullan, M. & A. Pomfret, Research on curriculum and instruction implementation. *Review of educational research*, 1977, vol. 47, no. 1.
- Grift, W. van de, Over kennisbelangen en de rollen van evaluatie-onderzoek. *Pedagogisch Tijdschrift*, 1981, 5, no. 8.
- Grift, W. van de & M. de Jong, *De externe evaluatie van het Landelijk Leerplaatsenproject*, deel II: Innovatie-inhoud en innovatiestrategie. Amsterdam: Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam, 1981.
- Grift, W. van de, G. de Vries & J. Baltzer, *De externe evaluatie van het Landelijk Leerplaatsenproject*, deel III: Het innovatieproces. Amsterdam: Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam, 1983.
- Groot, A. D. de, Psychologie. In: O. Noordenbos, *Wetenschap en leven*, de wetenschap als cultuurfactor in onze tijd. Haarlem: De erven F. Bohn N.V., 1959.
- Groot, A.D. de, Ontwikkelingslijnen in de Nederlandse onderwijsresearch. *Tijdschrift voor Onderwijsresearch*, 1976, 1, nr. 4, 145-160.
- Groot, A. D. de, Probleemgericht denken over onderwijs-evaluatie. In: P. Weeda, *Aspecten van leerplanevaluatie*. 's-Hertogenbosch: Malmberg, 1981.
- Gross, N., J. Giacuinta & M. Bernstein, *Implementing organizational innovations: a sociological analysis of planned educational change*. New York: Basic Books, 1971.
- Hall, G. E., S. F. Loucks, W. L. Rutherford & B. W. Newlove, Levels of use of the innovation: a framework for analyzing innovation adoption. *Journal of teacher education*, 1975, vol. 26, no. 1.
- Hall, G. E. & S. F. Loucks, A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal*, 1977, vol. 14, no. 3.
- Hall, G. E., *Innovation configurations: analyzing the adaptations of innovations*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978.
- Heck, S., S. M. Stiegelbauer, G. Hall & S. F. Loucks, *Measuring innovation configurations: procedures and applications*. Austin: The research and development center for teacher education, Texas, 1981.
- Hofstee, W. K. B., Evaluatie: een methodologische analyse. *Tijdschrift voor Onderwijsresearch*, 1982, 7, nr. 5.
- Hofstee, W. K. B., Evaluatie versus begeleiding. *Tijdschrift voor Onderwijsresearch*, 1983, 8, nr. 1.
- Hyman, H. H., C. R. Wright & T. K. Hopkins, *Applications of methods of evaluation*. Berkeley and Los Angeles: University of California press, 1962.
- Hyman, H. H. & C. Wright, Evaluating social action programs. In: P. Lazarsfeld, W. Sewell & H. Wilsensky (Eds.), *Uses of Sociology*. New York: 1967.
- Janssens, S., De waarde van begeleidingsinterventies in het kader van strategieën voor onderwijsinnovatie. *Pedagogisch Tijdschrift*, 1980, 5, nr. 4.
- Klauw, C. F. van der, *Evaluation of education: a review*. Rotterdam: Erasmusuniversiteit, 1979.
- Lagerweij, N. A. J., De veranderbaarheid van het onderwijs. In: J. A. van Kemenade (red.), *Onderwijs: bestel en beleid*. Groningen: Wolters Noordhoff, 1981.
- Leithwood, K. A. & D. J. Montgomery, Evaluating program evaluation. *Evaluation review*, 1980 vol. 4, no. 2.

- Leithwood, K. A. & D. J. Montgomery, A. framework for planned educational change: application to the assessment of program implementation. *Educational evaluation and policy analysis*, 1982, vol. 4, nr. 2, 157-167.
- Meijer, K., Naar een meta-evaluatiemodel voor formatieve evaluatie. In: P. Weeda, *Aspecten van leerplanevaluatie*. 's-Hertogenbosch: Malmberg, 1981.
- Mort, P. R., Studies in educational innovation from the institute of administrative research: an overview. In: M. B. Miles, *Innovation in education*. New York: Teacher's college press, 1964.
- Onderwijsonderzoek in opspraak, over de dilemma's van de onderwijsvernieuwing en de onderwijsevaluatie: Het Innovatieproject Amsterdam. *Info*, 1983, 14, nr. 2-3.
- Perre, C. van der, Configuraties als realiseringswijzen van een innovatie. In: R. van den Berg & R. Vandenberghe, *Onderwijsinnovatie in verschuivend perspectief*. Tilburg: Zwijzen, 1981.
- Provus, M., *Discrepancy evaluation*. Berkely: McCutchan Publishing corporation, 1971.
- Rogers, E. M. & F. F. Shoemaker, *Communication of innovations*, a cross cultural approach. New York: Free Press, 1971.
- Revicki, D. A., R. I. Rubin & G. B. Stuck, Models for measuring programme implementation: a review and critique. *Journal of curriculum studies*, 1981, vol. 13, nr. 1.
- Saris, W. E., *Introduction to the use of linear structural equation models in non-experimental research*. Amsterdam, z. j.
- Sandbergen, S., Testslimheid, de invloed van oefening en coaching op testcores. *Nederlands Tijdschrift voor de Psychologie*, 1968, deel XXIII, no. 8.
- Scriven, M., The methodology of evaluation. In: R. Tyler, R. Gagne & M. Scriven, *Perspectives of curriculum evaluation*. Chicago: Rand McNally, 1967.
- Slavenburg, J. H. (eindred.), *Het project onderwijs en sociaal milieu*. Tilburg: Zwijzen, 1978.
- Slavenburg, J. H., Bespreking van het eindverslag van het GEON-project. *Tijdschrift voor Onderwijsresearch*, 1982, 7, nr. 6
- Stake, R. E., The countenance of educational evaluation. *Teachers college record*, 1967, 68.
- Stokking, K. M. & A. K. de Vries, *Een luis in de pels*, eindverslag van het GEON-project. 's-Gravenhage: Staatsuitgeverij, 1981.
- Stokking, K. M., Evaluatie-onderzoek. In: P. G. Swanborn & L. Rademaker (red.), *Sociologische grondbegrippen*. Utrecht: Prisma, 1982.
- Stokking, K. M., Reactie op Hofstee: 'Evaluatie: een methodologische analyse'. *Tijdschrift voor Onderwijsresearch*, 1983, 8, nr. 1.
- Stufflebeam, D. L., Evaluation as enlightenment for decision making. In: W. H. Beatty (red.), *Improving educational assessment and an inventory of measures of affective behavior*. Washington, D. C., 1969.
- Stufflebeam, D. L., W. J. Gephart, E. G. Guba, R. L. Hammond, H. O. Marrison & M. M. Provus, *Educational evaluation and decision making*. Itasca, Illinois: Peacock Publishers Inc., 1971.
- SVO, *Evaluatie-onderzoek*, achtergronden en uitgangspunten voor beleidsvoering. 's-Gravenhage: SVO, 1980.
- Tyler, R. W., *Basic principles of curriculum and instruction*. Chicago: The university of Chicago press, 1949.
- Tyler, R. W., Outlook for the future. In: R. W. Tyler, *Educational evaluation: new roles, new means* (NSSE). Chicago: The university of Chicago press, 1969.
- Verbeeten, H. M. J. M., Ervaringen met een model voor formatieve evaluatie van kleuterschoolprogramma's. In: J. H. Slavenburg (eindred.), *Het project onderwijs en sociaal milieu*. Tilburg: Zwijzen, 1978.
- Viersen, D. J. van & P. F. A. Boersma, *Met het Mavoproject onderweg 2*, een onderzoek onder leerlingen. Nijmegen: ITS, 1980.
- Vries, A. de, *Integraal rapport van het instroomonderzoek middenschoolexperimenten 1981/1982*. Haren: RION, 1982.
- Wagenaar, J., N. van Kessel & A. de Vries, *Eindrapport project Beschrijving Beginsituatie Experimenten Middenschool II*, Een onderzoek naar de beginsituatie van scholen die als tweede generatie begonnen zijn aan het middenschoolexperiment. Nijmegen, Haren: ITS, RION, 1979.
- Weeda, P. (red.), *Aspecten van leerplanevaluatie*. 's-Hertogenbosch: Malmberg, 1981.
- Weeda, P., Cito en leerplanevaluatie. In: P. Weeda, *Aspecten van leerplanevaluatie*. 's-Hertogenbosch: Malmberg, 1981.
- Weiss, C. H., *Evaluation research*, New Jersey: Prentice Hall, 1972.
- Welch, W. W., Twenty years of science curriculum development: a look back. In: D. C. Berliner, *Review of research in education*. Itasca, Illinois: Peacock Publisher Inc., 1979.
- Zaltman, G. & R. Duncan, *Strategies for planned change*. New York: Wiley, 1977.

Curriculum vitae

W. van de Grift (1951) is als psycholoog verbonden aan het Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam. Hij werkte onder meer aan het onderzoek ter externe evaluatie van het Innovatieproject Amsterdam, aan het onderzoek ter externe evaluatie van het Landelijk Leerplaatsenproject en aan het Schoolgericht onderzoek Middenschool. Voorts voerde hij literatuuronderzoek uit op het gebied van evaluatie van onderwijsvernieuwingen. Momenteel werkt hij aan het construeren

van een instrument voor het meten van functionele taalvaardigheid en werkt hij aan de voorbereiding van het onderzoek ter externe evaluatie van het verkeerseducatieproject Schoolverkeer Moet Veiliger. Tevens is hij coördinator van de onderzoeksthema-groep Onderwijsvernieuwingen.

Adres: Stichting Centrum voor Onderwijs-onderzoek van de U.v.A., Keizersgracht 119, 1015 CJ Amsterdam

Manuscript aanvaard 27-10-'83