

Evalueren van Onderwijsinnovaties*

L. J. VAN DEN BOSCH

Onderzoeksgroep Participatie-Onderwijs, Utrecht

1. Het probleem

Algemeen gesteld zou men het onderwerp van de conferentie kunnen formuleren als: Hoe kunnen overheid, onderzoekers en onderwijsmensen hun verantwoordelijkheid voor het evalueren van onderwijsinnovaties gestalte geven?

In een paper en een inleiding werd door prof. R. Stake van de Universiteit van Illinois, die door CERI als expert voor dit projekt was aangetrokken, het onderwerp van de conferentie nader toegelicht. Stake konstateerde, dat de verantwoordelijkheid voor het onderwijs niet meer expliciet berust bij de direkt betrokkenen – de leerkracht, leerling en ouders – maar in toenemende mate wordt gedeeld met mensen en instanties die verder verwijderd zijn van het onderwijsleerproces. Allerlei representanten van de overheid en andere maatschappelijke instituties zijn mede verantwoordelijk voor en hebben bemoeienis met steeds meer aspecten van het onderwijssysteem.

Voor het nemen van beslissingen hebben deze instanties informatie nodig. Daarom is het nodig dat de evaluatie van vernieuwingsplannen niet alleen gebaseerd is op persoonlijke, subjektieve indrukken en fragmentarische evaluatie-activiteiten, maar tevens op systematisch en breed opgezet evaluatieonderzoek. Systematisch evaluatieonderzoek is een noodzakelijke aanvulling op de ervaringskennis van de praktikus en de theoretische inzichten van de wetenschapper.

Evaluatieonderzoek zal echter zelden gemakkelijke antwoorden geven op complexe problemen. Goed onderzoek zal leiden tot een verandering in de inzichten ten aanzien van het totale vernieuwingsprojekt en lang niet alle evaluatieonderzoekingen zijn goed opgezet.

Een belangrijke vraag is nu: Welke kondities zijn bepalend voor een goede start van evaluatieonderzoek?

* Een verslag van de OECD/CERI-konferentie 'New Approaches to Evaluation of Educational Programmes', gehouden te Luik van 2-5 december 1974.

Een belangrijk onderwerp in de conferentie was dan ook de kommunikatie tussen beslissers en evaluatieonderzoekers voor de aanvang van het onderzoek. Welke kwesties moeten besproken worden? Waarover moet overeenstemming worden bereikt? Welke procedure-afspraken moeten worden gemaakt? enz. In de conferentie werd vrij algemeen naar voren gebracht, dat de discussie tussen beslissers en onderzoekers en het maken van duidelijke afspraken van essentieel belang is, maar tevens dat deze kommunikatie tijdens het onderzoek op een bepaalde wijze moet worden voortgezet. Een evaluatieonderzoek, dat van enige betekenis wil zijn voor beslissers, heeft een complexe vraagstelling en flexibele opzet. Het is onmogelijk en ook niet gewenst dat voor de evaluatie van een complex innovatieprojekt een volledig uitgewerkt onderzoeksplan in de voorbereidingsfase wordt opgesteld. Stake formuleerde de volgende slogan:

'Get it all straight in the beginning but you can't'.

2. Toespitsing van het probleem

In de conferentie ging het in het bijzonder om breed opgezette, landelijke evaluatieonderzoekingen, die waarschijnlijk slechts eenmalig worden uitgevoerd en niet herhaald kunnen worden. Het doel van dergelijk onderzoek is in het algemeen informatie te verschaffen aan diegenen, die op een of andere wijze betrokken zijn bij het nemen van beslissingen ten aanzien van een complex en meestal landelijk innovatieprojekt.

Hoewel evaluatieprocedures voor de normale gang van zaken in het onderwijssysteem eveneens belangrijk zijn, viel het buiten het onderwerp van deze conferentie. De nadruk lag meer op onderzoek ten behoeve van een totaal-evaluatie van een complex en landelijk innovatieprojekt, dat meestal al in uitvoering is, dan op onderzoek ten behoeve van bepaalde topics zoals het evalueren van een nieuwe leesmethode of van selektiemethoden. Dergelijke topics zijn uiteraard ook belangrijke aspecten van

de eerste vorm van evaluatieonderzoek, maar als de aandacht primair hierop gericht is zullen meer traditionele methoden van onderwijsresearch eerder in aanmerking komen.

Deze toepassing van onderzoek als hulpmiddel voor de evaluatie van omvangrijke innovatieprojecten werd door de voorbereiders van de conferentie als 'program-evaluation' betiteld. Het woordje 'program' hierin kan misverstanden opwekken, hetgeen dan ook op de conferentie gebeurde. In het vervolg van dit verslag hanteren wij de term 'overall-evaluatieonderzoek', d.w.z. onderzoek dat bedoeld is om beslissingen die gebaseerd zijn op een (zo systematisch mogelijke) evaluatie van een complex, meestal landelijk innovatieproject mogelijk te maken.¹

Uiteraard is de evaluatie van een innovatieproject niet alleen gebaseerd op onderzoek. Tot nu toe was deze evaluatie meestal juist niet gebaseerd op onderzoek. Iedere betrokken leerkracht en ambtenaar evalueert. Terecht stelde Stake, dat informele en subjectieve evaluatie niet minder belangrijk dient te worden door de uitvoering van formeel overall-evaluatieonderzoek. Het onderzoek kan informatie verschaffen aan beslissers, die in een positie verkeren waarin ze geen persoonlijke observaties kunnen doen of die niet over middelen beschikken om de informele informatie afkomstig van vele scholen te combineren. Overall-evaluatieonderzoek verschaft komplementaire informatie over grote projecten en deze informatie onderscheidt zich van de informele informatie doordat ze nauwkeuriger, objectiever en meer valide is.

Het belang van overall-evaluatieonderzoek maakt evenmin de noodzaak van 'basic-research' op onderwijskundige vragen minder belangrijk. Met name een multi-disciplinaire benadering van deze vragen dient te worden gestimuleerd. Overall-evaluatieonderzoek zal echter vooral van betekenis zijn voor de beslissers in het onderwijssysteem - met name beleidsinstanties -, indien het in de eerste plaats gericht is op de vragen van deze beslissers ten aanzien van een innovatieproject in plaats van op een verdere ontwikkeling van de wetenschappelijke theorievorming over het onderwijs.

3. De opzet van de conferentie

De conferentie had het karakter van een workshop. Inhoudelijk was de conferentie voorgestruktuurd door middel van:

- enkele papers, waarin een aantal aspecten van het probleem werden verkend,

- een beschrijving van een fictieve case met een lijst van topics om de case te behandelen,
- het bijeenbrengen van een 7-tal beschrijvingen van evaluatieonderzoeksprojecten.

Met name door de fictieve case werd de relatie tussen 'administrators' of representanten van beleidsinstanties en onderzoekers centraal gesteld. Dit sloot vrij goed aan op de samenstelling van de deelnemersgroep, die voor iets meer dan de helft uit 'administrators' bestond en voor iets minder dan de helft uit mensen die een functie in de researchwereld bekleden. Er waren 36 deelnemers, afkomstig uit 14 landen.

Organisatorisch was de conferentie weinig gestructureerd. Na een toelichting van Stake op zijn papers werd er gedurende 2 dagen in 4 subgroepen aan de problematiek gewerkt. De subgroep-bijeenkomsten werden alleen onderbroken door een korte plenaire discussie met een forum over de 'responsive' versus 'metric approach' (zie pag. 132).

De subgroepen vatten de resultaten van hun werk samen in een kort verslag. Op de laatste dag gingen de 4 subgroepen paarsgewijs in discussie en reageerden op elkaars verslagen. De conferentie werd besloten met een plenaire evaluatiezitting.

4. Huidige situatie ten aanzien van overall-evaluatieonderzoek

Op basis van een aantal gesprekken met beleidsmensen en onderzoekers, van de bestudering van een aantal cases en van eigen ervaringen, schetste Stake de huidige situatie ten aanzien van overall-evaluatieonderzoek.

4.1. De benadering van overall-evaluatieonderzoek door beleidsinstanties

Over het algemeen zijn het vooral overheidsfunctionarissen die behoefte hebben aan overall-evaluatieonderzoek. Zij nemen vaak het initiatief, financieren de projecten en zijn vaak optimistischer over de bruikbaarheid van de resultaten dan de onderzoekers zelf. Dit is ook niet verwonderlijk. Zij staan voor de taak de beslissingen te nemen over de besteding van een beperkt budget, terwijl meestal veel te weinig informatie beschikbaar is. Zij hopen dat evaluatieonderzoek hen inzicht verschaft in kosten en uitkomsten van een project. In feite is hun steun aan een evaluatieonderzoek meer gebaseerd op hoop dan op positieve ervaringen.

Representanten van de overheid hebben meestal veel aandacht voor de communicatie met onder-

zoekers. Zij weten hoe moeilijk het is anderen te informeren en hebben een sterk geloof in de waarde van gedetailleerde schriftelijke rapportage. Naast de voordelen noemt Stake echter ook een aantal nadelen van frekwente schriftelijke rapportage. Het kost veel tijd, is altijd een vereenvoudigde weergave van de werkelijke situatie van het onderzoek en leidt soms tot verkeerde interpretaties. Beleidsfunctionarissen hebben de neiging deze nadelen te negeren.

Bovendien vormen de vraag om evaluatieonderzoek en de aanwezigheid van verslagen op zichzelf bewijzen, dat de beleidsfunctionaris zijn werk doet. Onbruikbare of misleidende verslagen leiden zelden tot veroordeling van degene die om de verslagen verzocht heeft. Soms bestaat de indruk, dat beleidsinstanties meer uit zijn op de erkenning dat zij inspelen op de behoefte aan systematische evaluatie, dan op het vinden van een adequaat antwoord op deze behoefte.

Een ander belangrijk aspect in de benadering van overall-evaluatieonderzoek door beleidsinstanties is, dat zij de neiging hebben – bewust of onbewust – hun vragen zodanig te stellen dat de verantwoordelijkheid voor het innovatieproject meer gecentraliseerd wordt en verder komt af te liggen van de scholen en projecten in het veld.

4.2. De benadering van overall-evaluatieonderzoek door onderwijzers

In tegenstelling tot beleidsfunctionarissen staan – volgens Stake – leerkrachten over het algemeen vrij afwijzend tegenover formeel evaluatieonderzoek. Hoewel zij erkennen dat de overheid in toenemende mate betrokken is bij kwesties binnen het onderwijs, beschouwen ze dit als een ongunstige ontwikkeling. Leerkrachten hebben het idee, dat iedere poging om een project te evalueren zal leiden tot een verkeerd beeld van tenminste een deel van het project en de kansen dat het project door het onderzoek negatief wordt beïnvloed is veel groter dan de kansen dat het project verder geholpen wordt.

Er is een zekere angst voor formeel evaluatieonderzoek en deze angst is ook niet geheel ongegrond.

Leerkrachten zijn niet tegen evaluatie als zodanig – beoordeling vinden de meesten immers een wezenlijk onderdeel van hun werk – maar ze geloven niet, dat evaluatieonderzoek de kwaliteit van het beslissingsproces zal verbeteren. Zij slagen er niet in de relevantie voor hun werk te zien van onderzoeken, die – zoals zij meestal hebben ervaren – uitgaan van het primaat van objectiviteit en systematiseren.

Opvallend is dat leerkrachten weinig of geen bezwaar naar voren brengen tegen nationale en inter-

nationale onderzoeken. Hoewel er een toenemende weerstand is te bespeuren, is tot op heden de medewerking van leerkrachten bij dergelijke onderzoeken geen grote belemmering geweest.

Officieel zijn de meeste organisaties van leerkrachten voorstanders van evaluatieonderzoek, waarbij ze meestal wel opmerken dat de primaire verantwoordelijkheid voor evaluatie bij het onderwijs zelf dient te berusten. In een aantal gevallen participeren lerarenorganisaties in de financiering van evaluatieonderzoek.

Samenvattend is Stake van mening, dat in de meeste gevallen leerkrachten bereid zijn positief te reageren op goed gepresenteerde en redelijke verzoeken om medewerking aan evaluatieonderzoek ten aanzien van hun scholen, curricula en vorderingen van de leerlingen.

4.3. De benadering van overall-evaluatieonderzoek door onderzoekers

De meeste onderwijsresearchers willen graag meehelpen innovatieprojecten te evalueren. In deze groep leeft in sterke mate de overtuiging dat een rationele benadering van onderwijsprocessen en onderwijsbeleid zal resulteren in een verbetering van deze processen en in een beter beleid. Zij hebben niet veel vertrouwen in beslissingen die gebaseerd zijn op ervaringskennis van beleidsfunctionarissen. Het ideaal van menig onderwijsresearcher is een beslissingsproces, dat gebaseerd is op objectieve metingen met behulp van gevalideerde instrumenten.

De aanbevelingen van onderzoekers zijn nog zeer vaak gebaseerd op een benadering van het onderwijs uitsluitend via de bril van hun specialisme. Onderwijspsychologen geven prioriteit aan het meten van individuele kenmerken. Onderwijs sociologen dringen aan op rol-analyse en onderzoek van de maatschappijstructuur. Onderwijskundigen benadrukken een meer planmatige ontwikkeling van curricula. Volgens Stake is deze specialisatie in de onderwijsresearch begrijpelijk, maar geen goede basis voor de ontwikkeling van concepten en methoden voor overall-evaluatieonderzoek.

Ook de motivatie van onderzoekers is soms een belemmerende faktor voor een adequaat antwoord op de behoefte aan evaluatieonderzoek.

Het is geen vanzelfsprekende zaak, dat het leveren van bijdragen aan algemene theorievorming, stage-mogelijkheden creëren voor studenten, het produceren van een goed wetenschappelijk rapport, enz. ten allen tijde kan samengaan met een adequaat inspelen op de evaluatiebehoeften van de beslissers in een innovatieproject.

De meeste onderzoekers formuleren een aantal ideale criteria voor evaluatie-designs: Het moet rationeel gedetailleerd gepland en voorbereid zijn; het moet concepten operationaliseren; het moet beslissers feedback verschaffen; etc. Weinigen zullen deze criteria bestrijden, maar sommige onderzoekers zullen lagere normen hanteren teneinde andere kenmerken van het design te benadrukken en andere doelen te bereiken.

De meeste onderzoekers reageren op de behoefte aan overall-evaluatieonderzoek in de verwachting dat de bekende methoden en concepten opnieuw gebruikt kunnen worden. Men heeft wel kennis genomen van nieuwe modellen en termen, maar deze nieuwe informatie komt nog nauwelijks tot uitdrukking in het concrete onderzoek.

Een probleem is dat verschillende mensen verschillende verwachtingen hebben ten aanzien van goed evaluatieonderzoek. De beleidsfunctionaris verwacht informatie, die hem helpt meer verantwoord te beslissen welke projecten (en in welke mate) hij zal subsidiëren. De leerkracht verwacht informatie waardoor hij een beter inzicht krijgt in een bepaalde onderwijsleersituatie. De onderzoeker is gericht op een algemeen inzicht in het onderwijsleerproces. Eén onderzoeksproject kan onmogelijk al deze verschillende behoeften bevredigen. Nodig is dan ook een klassifikatie van de verschillende doelen van evaluatieonderzoek in relatie met de klassifikatie van verschillende strategieën voor evaluatieonderzoek.

5. Een klassifikatie van evaluatiestrategieën

Desituaties waarin behoefte is aan overall-evaluatieonderzoek kunnen zeer verschillend zijn: verschillen in informatiebehoeften, verschillen in normen en verwachtingen van beslissers, verschillen in stijl van onderzoekers, etc. Het is dan ook noodzakelijk dat voor verschillende situaties verschillende evaluatiestrategieën worden ontwikkeld.

Stake, ontwikkelde een typologie, die enerzijds gebaseerd is op een combinatie van de dimensies van Scriven (1967) en Worthen en Sanders (1972) en anderzijds op een analyse van een groot aantal projecten.

In zijn typologie onderscheidt hij 9 evaluatiestrategieën. Sommige van de strategieën overlappen elkaar en in de praktijk zullen meerdere strategieën meestal in combinatie worden toegepast (zie schema Tabel 1).

'*Student gain by testing*'-strategie wordt vooral toegepast door toetspecialisten en onderwijspsychologen en is gericht op het meten van de vorderingen van de leerlingen door middel van objectieve toetsen.

Belangrijke elementen in deze strategie zijn o.a.: specificatie van de leerdoelen, het opsporen van de diskrepantie tussen de doelen en de bereikte leer-effecten en nagaan welke effecten alleen toegeschreven kunnen worden aan het onderwijsproces.

'*Criterion referenced tests*' worden gebruikt om na te gaan in welke mate de leerling een bepaalde taak beheerst of daarin vorderingen heeft gemaakt.

Kritici van deze strategie menen o.a., dat in deze strategie onvoldoende tot uiting komt de betekenis van het onderwijs voor de leerlingen en dat de strategie weinig bijdraagt tot het opsporen en verbeteren van tekorten in het onderwijsproces.

'*Institutional self-study by staff*' of 'self-survey' wordt vooral veel in Amerika toegepast waar de scholen een grote mate van autonomie bezitten. In deze strategie verzamelt de school zelf de gegevens, interpreteert deze en stelt verbeteringen voor. Deze strategie gaat uit van en bevestigt de status quo van bestaande normen.

De staf van de school investeert veel tijd in het onderzoek. Een groot voordeel is, dat de verantwoordelijkheid voor de oplossing van de problemen volledig blijft berusten bij degenen die er direct mee geconfronteerd worden.

'*Prestige panel*' of 'kommissie van wijze mannen' is een strategie, waarbij gezaghebbende personen met een hoge reputatie een kommissie vormen. Deze strategie wordt meestal gebruikt als een laatste redmiddel om uit een krisissituatie te komen.

'*Transaction-observation*'-strategie richt zich primair op de processen in het onderwijs. Dit in tegenstelling tot de 'student-testing'-strategie, die zich primair richt op de effecten van het onderwijs.

De transaktionele observatie beschrijft en analyseert de activiteiten in een innovatieproject en geeft speciale aandacht aan situationele variabelen. De vraagstellingen zijn eerder afgeleid van de praktische gang van zaken in een project dan van een theorie of doelformulering. De methoden, die in deze strategie worden gebruikt, zijn afkomstig uit de antropologie, de etnografie, de geschiedeniswetenschap en de journalistiek. Naast meer systematische meting van de processen vormt een – meer fenomenologische – beschrijving van de processen een belangrijk onderdeel van de strategie. Dit leidt er o.a. toe, dat

Tabel 1. 9 Evaluatiestrategieën

Strategie	Doel	Belangrijke elementen	Nadelen	Voordelen
'Student gain by testing' (meten van leerlingprestaties)	Het meten van de prestaties en voortdrijvingen van leerlingen	Doeluitspraken, analyse van de testcores, diskrepantie tussen doelen en gemeten effecten	Verschraling van de onderwijsoeuen, verwaarlozing van processen	Grote aandacht voor de voordeelingen van leerlingen
'Institutional self-study by staff' (self-survey)	Evaluatie en verbetering van het functioneren van de staf	Werkgroepen; formulering van de normen door de staf; discussie; stimuleert professionalisme	Bevordert in-group mentaliteit; negeert normen van de buitenwereld	Bevordert probleem-bewustzijn en verantwoordelijkheidsbesef van de staf
'Prestige panel' (kommissie van wijze mannen)	Oplossing van crises en het voortbestaan van een instituut	Kommissie van gezaghebbende personen, bezoeken, analyse van bestaande documenten, nota's, e.d.	Uitstel van actie, te groot vertrouwen in de intuïtie van de commissieleden	Verzameling en bundeling van de inzichten; uitspreken van oordelen
'Transaction observation' (transaktionale observatie)	Verschaffen van inzicht in activiteiten en normen	Aandacht voor processen, observatie, case studies, pluralisme	Te groot vertrouwen in subjektieve percepties, verwaarlozing van causale relaties	Verschaft een overall-beeld van een projekt, signaleert waarden-tegenstellingen
'Management analysis'	Bevordering van de rationaliteit in de dagelijkse beslissingen	Een lijst met opties, schattingen, feedbackprocedures, aandacht voor financiële aspecten en efficiency	Over-accentuering van efficiency, onderwaardering van impliciete normen	Feedback aan beslissers
'Instructional research'	Generaliseerbare theorie over het instructieproces	Gekontroleerde kondities, multi-variante analyse, hypothesevorming en -toetsing	Kunstmatige kondities, negeren van situationele en menselijke variaties	Ontwikkeling van nieuwe onderwijsmethoden
'Social policy analysis' (beleidsanalyse)	Helpen beleid te formuleren	Meten van sociale kondities en de implementatie van beleid	Verwaarlozing van onderwijskundige kwesties en details	Leidt tot beleidskeuzen, verheldert alternatieven
'Goal-free evaluation' (doelvrije evaluatie)	Het schatten van de effecten van een projekt	Negeert geformuleerde doelen, volgt een checklist	Overwaardering van documenten en rapportage	Gegevens over effecten met weinig voorstrukturering vanuit vastgestelde doelen
'Adversary evaluation' (polarisatiestrategie)	Het doen van een keuze uit twee alternatieven	Debatten tussen voor- en tegenstanders, kruisverhoren, jury	Personalistisch, oppervlakkig, tijdgebonden	Sterke informatie impact; toetsing van tegengestelde opvattingen aan elkaar

door middel van deze strategie meer inzicht ontstaat in de pluriformiteit van normen en waarden in het onderwijs.

'Instructional research'-strategie is enigszins verwand met de 'student-testing' benadering. Als onderwijsresearchers iets willen evalueren proberen zij een plan op te stellen, dat zo dicht mogelijk in de buurt komt van een experimenteel design. Zij streven er naar een bijdrage te leveren aan de algemene theorievorming, waarbij ze soms weinig aandacht hebben voor de vraag of het onderzoek nog relevant is voor de mensen die betrokken zijn bij dit innovatieproject.

Hewel deze benadering nog wijd verspreid is, neemt de kritiek toe op de vaak geringe feedback aan de direkt betrokkenen. Stake haalde het volgende citaat van John Nesbit aan:

'(For evaluating innovations) I think we can safely discard as inappropriate the original American model of precisely defined behavioural objectives, measurable outcomes, standardised tests and elaborate statistical analysis. This was summative external evaluation: summative in that it was applied at the end (or sometimes only at the end of the initial stage); external in that it was done by a group separate from the innovators, supposedly neutral but often unsympathetic. Instead, we need a continuous or ongoing evaluation, built into the innovative programme as an integral part of the team work. This is the model which was developed in the Schools Council 5-13 Project, and the Humanities Curriculum Project. The evaluator is an important member of the team, and he has one of the most difficult jobs to do. He is involved throughout in the planning (to ensure that evaluation is possible), during the programme (to ensure that relevant records are kept), and at the end of each stage (to provide feedback to the team on their strengths and weakness). It is a pattern of formative or responsive evaluation.'

Wanneer echter het doel van evaluatieonderzoek is generaliseerbare informatie te verschaffen, komt de 'instructional research' uiteraard wel in aanmerking.

'Management analysis' en 'social policy analysis' zijn nauw aan elkaar verwant. De verschillen hebben vooral betrekking op de urgentie van de informatie en de soort beslissers, waarvoor de informatie is bestemd.

Management-analyse tracht de 'leiding' van een innovatieproject te helpen bij de dagelijkse beslissingen.

Beleids-analyse is meer gericht op beleidsbeslissingen, die meestal betrekking hebben op een langer termijn. Het beslissingsproces is veelal gekomplieerder en de participanten hebben meer tot taak het beleid te formuleren dan het uit te voeren.

Sterk vereenvoudigd zou men kunnen stellen, dat management-analyse gericht is op ondersteuning van beleidsuitvoerders en beleidsanalyse op ondersteuning van degenen, die het beleid formuleren en vaststellen.

De twee strategieën hebben echter veel overeenkomsten. Beide maken gebruik van sociaal-wetenschappelijke concepten die gewoonlijk niet voorkomen in de onderwijsresearch. De management-strategie zal waarschijnlijk iets meer aandacht hebben voor economische kwesties en de beleidsstrategie voor politieke kwesties.

'Goal-free evaluation' en 'adversary evaluation'. Deze twee strategieën worden nog weinig toegepast, maar krijgen steeds meer aandacht. M. Scriven noemde de 'goal-free evaluation' in een discussie. Hij benadrukte een onafhankelijke positie van de evaluator (of onafhankelijke beoordelaar) ten opzichte van de staf. Hij meende zelfs, dat de evaluator van tevoren de doelen van de staf niet mag kennen. De evaluator is gericht op het identificeren van de doelen die in de praktijk worden nagestreefd en hanteert een groot aantal indicaties om na te gaan of er resultaten bereikt worden. Ten aanzien van de indicaties is de strategie dus niet 'goal-free'. Er wordt gebruik gemaakt van een sterk gestructureerde checklist.

In de 'adversary approach' of 'polarisatie-strategie' wordt de gang van zaken bij de rechtbank als voorbeeld genomen. Het evaluatieproces wordt opgezet langs twee lijnen: een lijn die er op gericht is de tekortkomingen van het innovatieproject te tonen en een lijn, die er op gericht is de waarde van het project te tonen. In navolging van de procedure bij een rechtbank vinden er 'kruisverhoren' plaats.

Het grote voordeel van deze strategie is, dat de bestede tijd in hoge mate gebruikt wordt om tot een beslissing te komen.

In de discussies op de conferentie kwam naar voren, dat de verschillende opvattingen van de deelnemers over evaluatiestrategieën schematisch weergegeven kunnen worden in twee tegengestelde benaderingen van overall-evaluatieonderzoek, n.l.:

- een 'responsive approach' of - vrij vertaald - een voorkeur voor meer kwalitatieve onderzoeksmethoden,
- een 'metric approach' of een voorkeur voor meer kwantitatieve onderzoeksmethoden.

De onderzoeker in een meer *kwalitatieve* benadering vergaart, selekteert, combineert, interpreteert en rapporteert vooral fenomenologische gegevens en in veel mindere mate meet- en beoordelingsresultaten. Hij legt de nadruk op procescriteria. De informatie die de beslisser ontvangt, bestaat daarvoor hoofdzakelijk uit een geordende weergave van door anderen opgedane ervaringen en de resultaten van een zorgvuldige reflexie op het project. Het grote voordeel van deze benadering is, dat de vraagstellingen van de onderzoeker in grote mate overeenkomen met de vragen van de beslissers. Daarnaast zijn de rapporten veelal minder technisch, beknopter en meer leesbaar voor niet-onderzoekers.

De onderzoeker in een meer *kwantitatieve* benadering vergaart, selekteert, combineert, interpreteert en rapporteert vooral kwantitatieve – statistische – gegevens of meetresultaten en in veel mindere mate fenomenologische gegevens. Hij legt de nadruk op produktcriteria. De metrische onderzoeker verschaft de beslisser informatie, die hoofdzakelijk bestaat uit relaties tussen bepaalde aspecten en verklaringen voor bepaalde effecten. Het voordeel van deze benadering is, dat de gegevens 'harder' zijn en met meer zekerheid gebracht kunnen worden. Het grote nadeel is, dat de vraag van de beslisser door de onderzoeker is gereduceerd tot enkele zeer specifieke vragen, die slechts betrekking hebben op details van het project. Hij kan alleen ingaan op vragen, waarvoor hij meetinstrumenten beschikbaar heeft. De rapporten, waarin een dergelijke benadering resulteert, staan door de consciëntieuse instelling van de onderzoeker soms vol met statistisch materiaal, zijn dan zeer omvangrijk en moeilijk toegankelijk voor niet-onderzoekers. Bovendien is dit onderzoek zeer tijdrovend en heeft het soms een stabiliserende werking op de dynamiek in een innovatieproject.

Als typering zou men kunnen stellen:

De metrische evaluatie-benadering evalueert de kwaliteit van de informatie, die aan de beslisser wordt verschaft.

De kwalitatieve evaluatie-benadering evalueert de kwaliteit van het inzicht of de verbetering van het inzicht of de kwaliteit van de beslissing.

De deelnemers aan de discussie waren het er in het algemeen over eens, dat iedere benadering zijn eigen waarde heeft en een keuze bepaald wordt door het te evalueren project. In relatie tot overall-evaluatieonderzoek toonde een gedeelte van de deelnemers een duidelijke voorkeur voor de 'responsive approach' of kwalitatieve strategieën. Het was onze indruk, dat deze groep vooral bestond uit de be-

leidsfunktionarissen en onderzoekers, die betrokken zijn bij op landelijk nivo opgezette innovatieprojecten. Een duidelijke afwijzing van de 'responsive approach' werd niet uitgesproken. De discussie ging vooral over de mate, waarin binnen deze meer kwalitatieve benadering gebruik gemaakt kan of moet worden van metrische methoden.

6. *Het overleg vóór het begin van een overall-evaluatieonderzoek*

Het overleg tussen de opdrachtgevers en de onderzoekers kan zowel te specifiek als te weinig specifiek gericht zijn. Wordt er teveel vastgelegd dan zal de onderzoeker niet in staat zijn met zijn onderzoek in te spelen op onverwachte moeilijkheden en mogelijkheden.

Vindt er in het overleg een zeer gedetailleerde bespreking plaats van het innovatieproject, het onderzoeksplan en de moeilijkheden die zouden kunnen ontstaan, dan is de kans groot dat er te weinig aandacht wordt besteed aan de grote beleidsvragen. Bovendien hebben de overlegpartners dan de neiging zichzelf in sterke mate te beschermen door middel van risicobeperkende clausules, waardoor de relevantie van het onderzoek kan worden belemmerd.

Is het overleg echter te weinig specifiek – hetgeen meestal het geval is – dan bestaat er slechts een geringe kans dat het evaluatieonderzoek zal ingaan op de vragen van de beslissers en dat het een positieve invloed zal hebben op het beslissingsproces.

Aspecten die in het overleg aan de orde dienen te komen zijn o.a.:

a. Doelen.

In het overleg dient ruime aandacht geschonken te worden aan de vragen, die bij de participanten in het beslissingsproces leven. Deze vragen dienen geïnventariseerd en goed begrepen te worden door de onderzoeker. Bij de belangrijkste vragen zal duidelijkheid moeten ontstaan over de informatiebronnen voor het evaluatieonderzoek en de resultaten die men van het onderzoek verwacht.

Een en ander betekent niet dat goed geformuleerde doelen persé noodzakelijk zijn of dat deze formuleringen een garantie vormen voor een juiste opzet van het onderzoek. Het belangrijkste is, dat de overlegpartners het gevoel hebben dat ze voldoende weten over de motieven voor een evaluatieonderzoek.

b. De lezers van het onderzoeksrapport.

De onderzoekers dienen te weten aan welke perso-

nen, groepen en instanties de opdrachtgever het rapport ter beschikking wil stellen. De onderzoeker zal ook moeten nagaan of daarnaast nog anderen een rol (kunnen) spelen in de beslissing die volgt op de rapportage van het onderzoek. Zij dienen op de hoogte te zijn van de vragen, die deze groepen het meest relevant vinden, van de normen die gehanteerd worden en van de wijze waarop de informatie het beste gepresenteerd kan worden.

Uiteraard zijn er binnen deze lezersgroepen ook belangrijke verschillen in relatie tot het object van evaluatie. De besteding van een gedeelte van de beschikbare mankracht en geld aan een bestudering van de lezersgroepen is vaak aan te bevelen.

De wijze, waarop de opdrachtgevers deze lezersgroepen percipiëren is een belangrijk aspect.

Daarnaast hebben de onderzoekers ook hun eigen achterban. Zij willen hun rapport publiceren in vakbladen, zij willen met hun projecten een goede indruk maken in de wereld van onderwijsresearch, etc. Ook dit dient in het overleg aan de orde te komen. Over de mate en wijze van disseminatie van de informatie dienen duidelijke afspraken te worden gemaakt.

c. De evaluatiestrategie.

Zowel de onderzoeker als de opdrachtgever zullen enige ideeën hebben over een adequate evaluatiestrategie. De discussie dient ertoe te leiden, dat de overlegpartners zich achteraf niet misleid voelen over de verwachtingen van de anderen. Een goede suggestie voor deze kwestie was, dat de overlegpartners voorbeelden noemen van goede en slechte evaluatieonderzoekingen.

Hoewel de onderzoekers in deze kwestie het laatste woord zullen hebben, kan de ervaring van de opdrachtgever in vroegere besprekingen met andere onderzoekers een belangrijke inbreng vormen.

Een opvallend verschijnsel voor beleidsfunctionarissen is, dat bekende onderzoekers zelden essentieel veranderen van werkwijze. De keuze van een onderzoeker is tegelijkertijd de keuze voor een bepaalde evaluatiestrategie. Flexibiliteit brengt men meestal alleen in zijn benadering door een onderzoeker met een andere stijl in het team op te nemen of een gedeelte van het onderzoek uit te besteden.

d. Vertrouwelijkheid.

Tijdens het voorbereidend overleg dienen duidelijke afspraken gemaakt te worden over de toegankelijkheid van de gegevens, voor welke doeleinden gegevens gebruikt worden en hoe lang en door wie ze bewaard blijven. Het blijkt dat er nogal eens moeilijkheden rijzen tussen opdrachtgevers en onder-

zoekers over de rapportage van als vertrouwelijk beschouwde informatie. Meestal is de onderzoeker dan ingegaan op kwesties, waarop niet van te voren is geanticipeerd. De gegevens worden door een aantal betrokkenen bij het innovatieproject wel als relevant beschouwd maar niet geschikt voor iedereen. Soms – als de gegevens zeer negatief zijn – voelt de onderzoeker zich moreel verplicht ze alleen aan de betreffende groepen te rapporteren, zelfs als afgesproken is dat niets gerapporteerd wordt zonder toestemming van de opdrachtgever of een bepaalde commissie. Dergelijke moeilijkheden zijn nooit geheel te voorkomen, maar het is aan te bevelen bij het voorbereidend overleg aandacht te schenken aan de mogelijkheden dat ze ontstaan.

e. Continuëring van het overleg.

Belangrijk is dat tijdens het voorbereidend overleg goede afspraken worden gemaakt voor regelmatige communicatie tijdens het onderzoek. Het is belangrijk, dat het mogelijk is op beslissingen terug te komen als men betere beslissingen kan nemen. Continuëring van het overleg is ook noodzakelijk, doordat overall-evaluatieonderzoek start met een flexibele opzet en de betrokkenen kunnen participeren in beslissingen die tijdens het onderzoek moeten worden genomen. Misverstanden kunnen voorkomen en bestreden worden door middel van communicatie, empathie en een bereidheid de juistheid van vroegere beslissingen of opvattingen te evalueren.

Hieronder volgt een lijst van topics, die prof. Stake tijdens de conferentie heeft opgesteld en die bedoeld is als hulpmiddel voor het overleg tussen beleidsfunctionarissen en onderzoekers.

*Checklist for negotiating an agreement to evaluate an educational program:

1. Do the parties to this negotiation know each other? What more do they need to find out? Who wants an evaluation study? Would those not participating (e.g., program developers, teachers, students) have added an important perspective?
2. What program is it that is to be evaluated? Whose program is it? What is its setting? its history? its purposes? its scope? How has it been evaluated before?
3. Why is there to be an evaluation study? What is it expected to produce? What should it accomplish (e.g., recommendations, authoritative judgements, explanations, points of view)?

4. Who are the audiences for the evaluation findings? Will different audiences (e.g. parents, technologists, members of parliament) have different background experience and different information needs?
5. What do the people who are most closely involved with the program see as its major issues or problems? What issues do other people see? How do all these relate to the major issues facing education elsewhere?
6. What resources are available for the conduct of this study? What cost estimates can be made (e.g., in money, staff, time, program disruption)?
7. What is the work history and working style of the prospective evaluators (i.e., the persons, team, or agency)? Do they have a portfolio of reports and artifacts from completed studies?
8. Why would the evaluators be interested in doing this study? What is there in it for them? Who else would they like to have helping them with it?
9. What will be the primary sources of data? What arrangements would be necessary to gain access to these sources? Are rules of access needed?
10. During the course of the evaluation study where and how would the data be kept? What would be the rules of access to these data (e.g., to participants, sponsors, newspaper reporters)?
11. What would be a suitable plan for reporting the findings? informal feedback? progress reports? final presentations? Are the evaluators free to publish findings in professional journals? What checks will be made on the effectiveness of the evaluation feedback?
12. How will further arrangements be negotiated after the study begins? What will be the response to unexpected changes in program? What misunderstandings may arise between the sponsors of the study and the evaluators? How will conflict be resolved?
13. What more needs to be said about the purposes and expectations for the evaluation study?

7. De kosten voor overall-evaluatieonderzoek

Evaluatieonderzoek is een relatief kostbare aangelegenheid, vooral vanwege de personeelskosten (ca 80%). Een veel voorkomende procedure o.a. in de Verenigde Staten is een bepaald percentage van de kosten van een innovatieproject te reserveren voor onderzoek. Bij grote projecten die volgens konven-

tionele methoden – summatieve evaluatie – worden geëvalueerd is dit meestal 2% en bij kleinere projecten 5%.

De kosten van formatieve evaluatie zijn meer afhankelijk van de problemen die tijdens het ontwikkelingswerk rijzen, maar zullen veelal hoger liggen dan bij summatieve evaluatie. Stake noemt een voorbeeld in de Verenigde Staten, waarin research-instituten niet meer dan 5% van de totale kosten van een innovatieproject op realistische wijze konden besteden. Het begroten van overall-evaluatieonderzoek is echter een moeilijk probleem.

De hoogte van het budget is geen goede indikator voor de kwaliteit van het onderzoek. Anderzijds kan het niet verstrekken van een realistisch krediet of het subsidiëren van een project voor korte perioden ertoe leiden dat de onderzoekers volledig gekoncentreerd raken op het opstellen van en onderhandelen over nieuwe subsidieaanvragen. Het onderzoek zelf kan dan volledig aan zijn doel voorbij schieten, bijv. doordat de onderzoekers evaluatiemethoden kiezen die algemeen geaccepteerd worden maar weinig relevante informatie opleveren voor de beslissers in het betreffende innovatieproject. De betekenis van menig onderzoek wordt aanzienlijk belemmerd door bureaucratische werkwijzen van subsidiënten.

Stake noemde een aantal variabelen bij het begroten van overall-evaluatieonderzoek:

- a. Bij een grote populatie is het werken met een steekproef goedkoper dan het enquêteren van iedere leerkracht of leerling. Bij een kleine populatie is een steekproefbenadering echter vaak duurder.
- b. Wanneer men bij het bepalen van de omvang van de steekproef uitgaat van de criteria van specialisten is de stijging van de kosten voor het verzamelen van data onevenredig groot in verhouding tot de omvang die noodzakelijk is om adequaat te reageren op de evaluatievragen.
- c. Het toepassen van bestaande instrumenten – ook al zijn deze niet geheel relevant – verdient de voorkeur boven het konstruëren en valideren van nieuwe instrumenten, gezien het verschil in kosten.
- d. Het gebruik van gestandaardiseerde instrumenten voor een klein aantal observaties is geldverspilling. Niet-gestandaardiseerde observatie-instrumenten of vragenlijsten zullen waarschijnlijk meer informatie verschaffen en zijn veel goedkoper bij kleine populaties.
- e. Komputers zijn onmisbaar voor analyse van een grote hoeveelheid gegevens. In vele evaluatieprojecten wordt er echter te veel en te snel aan

- methoden gedacht, waarbij de komputertechniek kan worden ingeschakeld. Deze kosten worden meestal snel geaccepteerd. Vooral in kleinere evaluatieprojecten dient de inschakeling van de komputertechniek kritisch te worden bekeken.
- f. Publicatie- en disseminatiekosten worden meestal veel te laag begroot.
 - g. Overheadkosten (dokumentatie, consultatie van deskundigen, etc.) zijn bij overall-evaluatieonderzoek even noodzakelijk als bij meer conventionele research.
8. *Diskussiepunten met betrekking tot overall-evaluatieonderzoek.*
- Door de voorbereiders van de conferentie was een fictieve case opgesteld. Als hulpmiddel voor de discussie werd een 'checklist' van discussiepunten uitgereikt. Hoewel deze lijst beslist niet volledig is, geeft ze toch een aardig overzicht van de vragen die aan de orde dienen te komen bij de voorbereiding van een overall-evaluatieonderzoeksproject.
- 8.1. De beslissingen, waarvoor evaluatieve informatie nodig is, de aard van deze informatie en de doelen waarvoor het kan worden gebruikt.
 - 8.1.1. Wat is het object van de beslissingen?
 - a. Gaat het om beslissingen met een sterk politiek karakter, die betrekking hebben op belangrijke veranderingen in het onderwijssysteem (bijv. invoering middenschool), maar die tevens zullen leiden tot een groot aantal veranderingen in de onderwijspraktijk?
 - b. Of gaat het om beslissingen, waarbij onderwijskundige argumenten de belangrijkste rol spelen, zoals invoering van een nieuw curriculum wetenschap, invoering van bijscholingskursus, etc.?
 - 8.1.2. Voor welk doel wil men de informatie gebruiken?
 - a. Voor het nemen van beslissingen ten aanzien van het al of niet uitbreiden en continueren van een bestaande experimentele situatie?
 - b. Voor het verkrijgen van meer acceptatie van een reeds ingevoerde vernieuwing?
 - c. Voor het nemen van beslissingen t.a.v. de modificatie of bijstelling van een reeds bestaande experimentele situatie of vernieuwing?
 - d. Voor het versterken van de opvattingen dat er verandering nodig is t.a.v. bepaalde traditionele situaties in het onderwijssysteem?
 - e. Voor het nemen van beslissingen die de verdere implementatie of verspreiding van een experiment effectiever doet verlopen?
 - 8.1.3. Welke soort informatie is noodzakelijk?
 - a. Nauwkeurige meting van leerlingprestaties?
 - b. Inzicht in de effecten op de attitudes van leerlingen?
 - c. Inzicht in de samenhang van factoren die een leerproces beïnvloeden?
 - d. Opvattingen en ervaringen van leerlingen?
 - e. Indirekte indicaties voor de reacties van leerlingen (absentisme, e.d.)?
 - f. De bij b. en e. genoemde informatie in relatie tot veranderingen in de context van de onderwijsleersituatie (leraren, schoolorganisaties)?
 - g. Als f., maar in relatie tot buitenschoolse factoren (beheersstructuur, ouders, maatschappij)?
 - 8.1.4. In welke mate wordt van het evaluatieonderzoek verwacht dat
 - a. het informatieverschaf op basis waarvan de beslissers hun eigen oordeel kunnen vormen (- en misschien rechtvaardigen)?
 - b. het - geheel op basis van de onderzoeksresultaten - tot een duidelijk oordeel komt?
 - c. het bewijst, dat een bepaalde beslissing juist zal zijn?
 - 8.1.5. In welke mate dient de informatie vanuit het onderzoek:
 - a. uitsluitend bestemd te zijn voor de opdrachtgevers, begeleidingscommissie en/of andere speciale groepen?
 - b. toegankelijk te zijn voor het grote publiek?
 - 8.1.6. Wat is de professionele verantwoordelijkheid van de onderzoeker in relatie tot 8.1.5.?
 - 8.1.7. In welke mate kan het onderzoeksproject de functie van hulpmiddel voor de beslissers combineren met de functies van middel tot een betere voorlichting van het publiek en ondersteuning van de begeleiders in het innovatieproject?
 - 8.1.8. Hoe gedifferentieerd of omvangrijk moet de informatie over de (mogelijke) effecten van beslissingen zijn?
 - effecten op leerlingen (cognitief en affectief),
 - effecten op leerkrachten,
 - effecten op vakbonden,
 - effecten op de interne schoolorganisatie,
 - effecten op de beheersstructuur,
 - effecten op ouders,
 - effecten op de maatschappij,
 - effecten op politici,
 - financiële consequenties,
 - effecten op de opleiding en bijscholing van leerkrachten.
 - 8.2. *De aanbesteding van evaluatieonderzoek.*
 - 8.2.1. Hoe komt men tot een juiste keuze voor de

leiding van het onderzoek? Welke gegevens zijn er nodig om te beslissen of de werkwijze van een bepaalde onderzoeker aansluit op de behoefte aan informatie van de beslissers?

Risico's:

- a. De onderzoeker is – of wordt gezien als – onvoldoende afhankelijk van de opdrachtgever en beslisser.
 - b. De normale werkwijze van de onderzoeker is niet geschikt voor overall-evaluatieonderzoek: hij gaat uit van het konsept dat alleen harde gegevens van leerlingprestaties valide zijn, terwijl de beslisser een breed geheel van informatie wenst.
 - c. De onderzoeker is geneigd evaluatieonderzoek te beschouwen als een stuk wetenschappelijk onderzoek met het oog op zijn eigen professionele doelen.
- 8.2.2. Verduidelijking van de doelen van het onderzoek, de mate van betrouwbaarheid, overdracht van de informatie aan andere groepen dan alleen de beslissers.
 - 8.2.3. Verduidelijking van de doelen en objecten van het innovatieproces, dat in het onderzoek geëvalueerd moet worden.
 - 8.2.4. Discussie over het karakter van de benodigde informatie en de betekenis van deze informatie in het beslissingsproces.
 - 8.2.5. De positie van de onderzoeker t.o.v. de beslissers en t.o.v. degenen die geëvalueerd worden. Hoe onafhankelijk moet hij zich gedragen t.o.v. de beslisser? Welke rechten heeft hij tot interventies in het innovatieproces?
 - 8.2.6. Welke evaluatiestrategie komt in aanmerking?
 - 8.2.7. Wat moet er in het onderzoeksplan vastgelegd worden en wat kan of dient open te blijven?
 - 8.2.8. Hoe wordt de kommunikatie tussen de onderzoeker en de beslissers tijdens het onderzoek georganiseerd?
 - 8.2.9. Welke toegang heeft de onderzoeker tot vertrouwelijke data-bronnen?
 - 8.2.10. Tijd- en kostenplanning.
- 8.3. Hoe zullen degenen die geëvalueerd worden, reageren op de onderzoekers?
 - 8.3.1. Reactie van leerkrachten: onzekerheid bij evaluatie van eigen functioneren.
 - 8.3.2. Het is mogelijk dat de onderzoekers gezien worden als representanten van de overheid.
 - 8.3.3. 'Onnatuurlijk' gedrag om een goed eindruk te maken.
 - 8.3.4. Enerzijds pogingen van leerkrachten en schooldirecties om de onderzoeker direct te

betrekken bij het programma en anderzijds de mogelijkheid, dat de onderzoekers geïdentificeerd worden met het projekt dat ze evalueren.

8.4. Welke evaluatiestrategie? (tabel 1)

- 8.5. Welke methoden voor het verzamelen van gegevens?
 - testen,
 - interviews,
 - vragenlijsten,
 - case-studies van delen van het projekt,
 - illustratie d.m.v. anekdoten waardoor belangrijke details duidelijk worden,
 - inhouds-analyse van rapporten, artikelen, enz.,
 - zelf-evaluatie door betrokkenen,
 - meer of minder gestructureerde observatie,
 - indirecte en pluriforme indicatoren, zoals absentisme van leerkrachten en leerlingen,
 - etc.
- 8.6. De presentatie van de informatie aan de beslissers en anderen.
 - 8.6.1. Wat is de waarde van een omvangrijk eindrapport met aan het eind een apart hoofdstuk 'Konklusies'?
 - 8.6.2. Wat is de waarde van de presentatie van statistische gegevens aan beslissers, die de werkelijke betekenis van deze gegevens slechts gedeeltelijk zullen begrijpen?
 - 8.6.3. In welke mate kan en mag er gebruik gemaakt worden van onderzoekstermen?
 - 8.6.4. Ontstaat er een konfliktituatie tussen het streven naar een 'wetenschappelijk' rapport en het streven naar een 'praktisch bruikbaar' rapport?
 - 8.6.5. Welke mogelijkheden zijn er voor de opstelling van verschillende rapporten met het oog op verschillende lezersgroepen?
 - 8.6.6. Zijn er mogelijkheden tot een meer actieve disseminatie van de informatie in de vorm van mondelinge toelichtingen, konferenties, workshops?
 - 8.6.7. Moet er tijdens het onderzoek kontinu gerapporteerd worden en hoe wordt dit dan georganiseerd?
 - 8.6.8. Zijn er mogelijkheden om audio-visuele hulpmiddelen te gebruiken bij de rapportage?
 - 8.6.9. Hoe moet de relatie zijn tussen de kosten van de rapportage en de totale kosten van het onderzoek?

9. De betekenis van de konferentie voor de nederlandse situatie.

9.1. Algemeen

De problematiek die op de konferentie werd besproken is in hoge mate relevant voor de nederlandse situatie. Wij danken hierbij met name aan een drietal ontwikkelingen.

In de eerste plaats heeft de overheid in grote mate bemoeienis met de gang van zaken binnen het onderwijs. Met name in Nederland neemt de overheid beslissingen over regelingen, die betrekking hebben op allerlei - vaak gedetailleerde - aspecten van het onderwijssysteem. De beheersstructuur heeft grote invloed op de processen binnen het onderwijs. In toenemende mate is de overheid zich ervan bewust geworden, dat haar verantwoordelijkheid niet beperkt dient te blijven tot een goed beheer van een normale gang van zaken of status quo in het onderwijs, maar tevens tot het stimuleren van verbeteringen in het onderwijssysteem. Door deze ontwikkeling vindt er thans een aksentverschuiving plaats in de rol van de overheid ten opzichte van het onderwijsveld. De overheid is nu niet meer alleen gericht op het bewaken van de kwaliteit van de bestaande situatie in het onderwijs, maar tevens op het initiëren van en het scheppen van condities voor belangrijke innovaties.

Samenhangend met het bovenstaande is in ons land vanuit het landelijk beslissingsnivo in een korte periode het initiatief genomen tot een aantal zeer complexe, ingrijpende en ambitieuze innovatieprojecten. We denken hierbij aan de innovatieprojecten participatie-onderwijs, basis-onderwijs, midden-school en open school. Ook door deze ontwikkelingen is de problematiek van overall-evaluatieonderzoek een interessant studie-objekt. De belangrijkste beslissingen ten aanzien van deze projecten worden genomen op centraal of landelijk nivo, waarbij politieke factoren een belangrijke rol spelen.

In de derde plaats blijkt met name uit de structurele voorwaarden, die ten aanzien van de 4 genoemde innovatieprojecten zijn uitgedacht, de intentie dat onderzoek een belangrijke rol dient te spelen in de ondersteuning van het beslissingsproces. Men hoopt, dat allerlei beslissingen in redelijke mate gebaseerd kunnen worden op de resultaten van evaluatieonderzoek. Het is op dit moment onzeker of de nederlandse onderwijsresearch een adequaat antwoord weet te geven op deze uitdaging. Wij hebben de indruk, dat dit met name bepaald zal worden door het tempo waarin en de wijze waarop men de belemmerende condities in de huidige situatie weet op te lossen. We denken hierbij o.a. aan zaken als

de organisatie van de onderwijsresearch, de financiering, de overlegprocedures, de kwalitatieve en kwantitatieve manpower, modellen voor de programmering van het onderzoek, de theorievorming met betrekking tot evaluatieonderzoek, enz.

De in dit verslag weergegeven ideeën lijken ons met name relevant te zijn voor de discussies over:

- strategieën voor evaluatieonderzoek t.a.v. de genoemde innovatieprojecten,
- de overlegstructuren tussen beleids(adviserende) instanties en onderzoekers,
- de overlegstructuur tussen onderzoekers en de subsidiënt van het onderzoek.

Teneinde de discussie over de problematiek van overall-evaluatieonderzoek in Nederland een impuls te geven en te verbreden lijkt ons een goed voorbereide, meer-daagse konferentie over deze problematiek aan te bevelen. Voor deze konferentie zouden dan met name uitgenodigd dienen te worden (potentiële) evaluatieonderzoekers, representanten van de landelijke beleidsinstanties en mensen uit het onderwijsveld.

9.2. Overall-evaluatieonderzoek t.a.v. het participatie-onderwijs

Sinds 1973 probeert een Onderzoeksgroep de onderzoeksfunctie in het projekt participatie-onderwijs concreet gestalte te geven. Een belangrijk onderdeel van de taak van deze Onderzoeksgroep komt overeen met wat wij in dit verslag 'overall-evaluatieonderzoek' genoemd hebben. Het projekt participatie-onderwijs is een landelijk en zeer complex projekt. Het was al 2 jaar in uitvoering, toen de eerste voorziening voor de voorbereiding van o.a. evaluatieonderzoek startte.

Ook in ons werk bij deze Onderzoeksgroep hebben wij ervaren, dat vooral beslissers op landelijk nivo behoefte hebben aan breed opgezet evaluatieonderzoek.

In de Onderzoeksgroep wordt geprobeerd strategieën te ontwikkelen, die inspelen op deze behoefte. Een probleem hierbij is, dat er nog weinig ervaring is opgedaan in 'overall-evaluatieonderzoek', waardoor o.m. een samenhangend geheel van concepten voor dit onderzoek nog moet worden ontwikkeld. Zo'n theoretisch framework is van belang als basis voor de discussie tussen onderzoekers onderling en tussen onderzoekers en beslissers en voor de opstelling van concrete onderzoeksplannen.

Bij ons pogen om 'overall-evaluatieonderzoek' in de sektor participatie-onderwijs te realiseren hebben wij de indruk gekregen, dat met name een tweetal factoren van belang is.

In de eerste plaats denken wij hierbij aan een

intensievere communicatie tussen landelijke beslissers en onderzoekers. De geringe bekendheid van de meeste beslissers met de mogelijkheden en het referentiekader van onderzoekers en de geringe bekendheid van onderzoekers met de problematiek van beslissers leidt over en weer tot irreële verwachtingen. Van de beslissers wordt bij een intensievere communicatie gevraagd, dat zij zich minder eenzijdig konsentren op actuele problemen en tijd vrij maken voor deze communicatie. Voor de onderzoekers impliceert deze communicatie, dat zij het overleg zien als een wezenlijk onderdeel van hun onderzoekswerk, niet bevreesd zijn enigszins geïdentificeerd te worden met 'het beleid' en de normen van hun professionele achterban ter discussie kunnen stellen. Gezien de structuur van en de verhoudingen in de wereld van de onderwijsresearch is er sprake van een zekere druk op onderzoekers zich te konformereren aan de heersende normen. Deze professionele 'afhankelijkheid' komt o.a. tot uiting in de beslissingsprocedures met betrekking tot subsidie-aanvragen voor onderzoek, waarbij gezaghebbende representanten uit de wereld van de onderwijsresearch een belangrijke rol spelen (S.V.O.-procedure).

De waardering voor 'meer kwantitatieve en metrische onderzoeksstrategieën' vormt een belemmering voor het creatieve denken over nieuwe, meer kwalitatieve strategieën, die uitgaan van complexere vraagstellingen.

Een tweede faktor bij het bevorderen van overall-evaluatieonderzoek is de aanwezigheid van een uitgewerkte landelijke planning op langere termijn. Doordat deze planning in de sector participatie-onderwijs op dit moment nog onvoldoende beschikbaar is, bestaat er veel onduidelijkheid over de doelen en objecten van het innovatieproces. Het is daardoor zeer moeilijk een kader – met name de keuze van criteria – te formuleren voor de beschrijving van het innovatieproces.

Resumerend zijn wij van mening, dat de start van overall-evaluatieonderzoek ten aanzien van het participatie-onderwijs een hoge prioriteit dient te hebben. Gezien het ontbreken van ervaring dient men de voorstellen voor dit onderzoek met een 'eksperimentele houding' te benaderen. We leren de waarde en de beperking van nieuwe benaderingen kennen door ze uit te proberen. Deskundigheid in overall-evaluatieonderzoek ontstaat door ervaring (ook 'mislukte' projecten kunnen op deze wijze waardevol zijn) en een gebrekkig evaluatieonderzoek is soms beter dan geen evaluatieonderzoek.

10. Enige relevante literatuur.

- Apple, Michal W. et al.: 'Educational Evaluation: analysis and responsibility', Berkeley, Mc.Cutchan, 1974.
- Campbell, Donald: 'Reforms as experiments', *American Psychologist*, 1964, 24, 409-429.
- Cohen, David K.: 'Politics and research-evaluation of large-scale-programs', *Review of Educational Research*, 1970, 40(2), 213-238.
- Coleman, James S.: 'Policy research in the social sciences', Morristown, New Jersey: General Learning Press, 1972.
- House, Ernest R.: 'Schoolevaluation, the politics and process', Berkeley, Mc.Cutchan, 1973.
- Mosteller, Frederick and Moynihan, David P. (eds): 'On equality of educational opportunity', New York, Vintage Books, 1972.
- Parlett, Malcolm and Hamilton, David: 'Evaluation as illumination: A new approach to the study of innovatory programs', University of Edinburgh, CERES, Occasional Paper 9, 1972.
- Popham, W. James: 'Evaluation in Education: current applications', Berkeley, Mc.Cutchan, 1974.
- Provus, Malcolm: 'Discrepancy evaluation', Berkeley, Mc.Cutchan, 1973.
- Riphey, Robert (ed.): 'Studies in transactional evaluation', New York, Mc.Cutchan, 1973.
- Scriven, M.: 'The Methodology of Evaluation' in 'Perspectives of curriculum evaluation' van R. W. Taylor, R. M. Gagné en M. Scriven, C.A.E.R.A. Monographs Series on Curriculum Evaluation, no. 1, Chicago, Rand Mc.Nally, 1967.
- Stake, Robert E.: 'The measuring of education', Berkeley, Mc.Cutchan, verschijnt in 1975.
- Stufflebeam, Daniel L., et al.: 'Educational Evaluation and Decision-making', Itasca, Illinois, Peacock, 1971.
- Tyler, Ralph W. (ed.): 'Educational Evaluation: new roles, new means', University of Chicago Press, 1969, 68th Yearbook, Part II, National Society for the Study of Education, 189-220.

Noot

1. Elders spraken wij van 'beleidsonderzoek' (L. J. van den Bosch en T. Wolthuis: 'Een theoretische verkenning van het begrip evaluatiestrategie', Utrecht, Onderzoeksgroep Participatie-Onderwijs, juni 1974). Hoewel wij deze term meer toepasselijk achten, wilden we in dit verslag niet te veel afwijken van de terminologie, die op de conferentie werd gebruikt.

Curriculum vitae

Drs. L. J. van den Bosch (geb. 1943) studeerde sociale pedagogiek te Utrecht. Was werkzaam als vormings-leider bij werkende jongeren, als leraar a.v.o. op een Havenvakhschool, als medewerker aan het Instituut voor Pedagogische en Andragogische Wetenschappen te Utrecht, begeleidde als medewerker van de Stichting Uitvoeringsgroep Streekcentra de ontwikkeling van experimenten participatie-onderwijs en is sinds 1973 coördinator van de Onderzoeksgroep Participatie-Onderwijs. Adres: Ravellaan 1, Utrecht.