

# Effecten van voorschoolse stimuleringsprogramma's: een review van reviews

H. Blok en P. P. M. Leseman

## Samenvatting

Bij de bestrijding van onderwijsachterstanden worden voorschoolse stimuleringsprogramma's in preventieve zin ingezet. In dit artikel wordt de vraag behandeld hoe sterk het effect van zulke programma's op de ontwikkeling van kinderen is. Ter beantwoording zijn tien reviews - verschenen tussen 1983 en 1993 - geanalyseerd. De reviews zijn op vier criteria beoordeeld: de selectie van de bronstudies, de codering van de kenmerken van de studies, de manier waarop de resultaten van de studies zijn geïntegreerd en de correspondentie tussen de gepresenteerde data en de getrokken conclusies. Acht van de tien reviews blijken ten aanzien van deze criteria ernstige tekorten te vertonen. De resterende twee reviews laten de volgende conclusies toe: a. na afloop van de interventie laten IQ-testen en schoolvorderingstoetsen een gemiddelde winst van ongeveer een halve sigma zien; b. de winst op het sociale domein bedraagt ongeveer een kwart sigma; c. in de meeste gevallen is de winst na één tot twee jaar niet meer aantoonbaar; d. effecten op de schoolloopbaan worden in slechts enkele studies aangetoond; e. er is weinig bekend over de consequenties van vormgevingskenmerken voor de effectiviteit van de programma's. Ten slotte, aangezien het meeste onderzoek tussen 1960 en 1980 is uitgevoerd bij Amerikaanse achterstandskinderen, is de generaliseerbaarheid van deze conclusies naar de huidige Nederlandse situatie waarschijnlijk beperkt.

## 1 Inleiding

De aandacht voor voorschoolse educatie is de afgelopen jaren duidelijk gegroeid. Eén van de factoren hierbij is het toenemend gebruik van voorzieningen als kinderdagverblijven en peuterspeelzalen. Een andere factor is de teleur-

stelling over de resultaten van de strijd tegen onderwijsachterstanden. Steeds meer wint het inzicht veld dat onderwijsachterstanden beter voorkomen dan verholpen kunnen worden. De veronderstelling is dat voorschools onderwijs ervoor kan zorgen dat kinderen die op hun vierde jaar de basisschool betreden, een min of meer gelijke start kunnen maken. Met name kinderen die thuis in een achterstandssituatie opgroeien, zouden van voorschools onderwijs profiteren.

Naar de effectiviteit van voorschoolse stimuleringsprogramma's is - in het buitenland - inmiddels heel wat onderzoek verricht. Gerekend vanaf de jaren zestig loopt het aantal afzonderlijke studies in de honderden. Het is voor individuele onderzoekers feitelijk onmogelijk om zo'n grote hoeveelheid onderzoek te overzien. Gelukkig zijn er overzichtstudies verschenen, die een goed startpunt kunnen vormen om tot een verantwoorde afweging te komen van de effecten van voorschools onderwijs.

Wij hebben de opdracht gekregen alle effectonderzoek, ná 1985 verschenen, te reviewen<sup>1</sup>. Dat gebeurt in twee stappen. In de eerste, voorbereidende stap - waarover we hier rapporteren - hebben we de reeds beschikbare, recente reviews geanalyseerd. In de tweede stap, waarover we in de nabije toekomst hopen te rapporteren, komen we aan onze eigenlijke opdracht toe, namelijk het reviewen van het na 1985 gepubliceerde effectonderzoek.

Wij hebben in totaal tien reviews - verschenen tussen 1983 en 1993 - geanalyseerd. Onze analyse is alleen gebaseerd op de reviews zelf. We zijn dus niet teruggegaan naar de bronstudies. In eerste instantie ging onze belangstelling vooral uit naar de conclusies die de auteurs aan hun reviews verbonden. Naarmate wij meer reviews verzamelden en steeds vaker constateerden dat de daarin vermelde conclusies op belangrijke punten divergeerden, kregen we meer belangstelling voor de kwaliteit

van de reviews. Dit artikel is dus meer dan alleen een beschrijvende samenvatting van eerdere reviews. Het is in belangrijke mate ook een kritische analyse van de wijze waarop de reviewers zich van hun taak hebben gekwetend.

## 2 Probleemstelling en opzet

### 2.1 Probleemstelling

In deze bijdrage onderscheiden we een primaire en een secundaire vraag. De primaire vraag is die naar de effecten van voorschoolse stimuleringsprogramma's. De secundaire vraag betreft de kwaliteit van de sinds 1980 verschenen reviews over dit onderwerp. Beide vragen worden toegelicht.

De *primaire* vraag luidt: 'Hoe sterk is het effect van voorschoolse stimuleringsprogramma's op de ontwikkeling van kinderen?' Bij de uitwerking van deze vraag hebben we de volgende specificaties aangebracht. Met de term stimuleringsprogramma's wordt op programma's gedoeld die expliciet tot doel hebben de ontwikkeling van kinderen te stimuleren. Het gaat hier om programma's die een uitgesproken educatieve doelstelling hebben. Zulke stimuleringsprogramma's onderscheiden zich van programma's die als hoofddoel hebben, kinderen een opvang te bieden voor de momenten dat ouders geen zorg willen of kunnen bieden. Zulke opvangprogramma's laten we buiten beschouwing, evenals programma's die alleen op ouders gericht zijn, bijvoorbeeld ter ondersteuning van hun opvoedingsgedrag.

Het merendeel van de stimuleringsprogramma's is 'center-based', hetgeen impliceert dat de kinderen het programma buiten hun eigen huis volgen. In sommige gevallen bevatten programma's naast een 'center-based' component ook een 'home-based' component, waarin bijvoorbeeld zaken als thuisbezoeken plaatsvinden.

De kwalificatie 'voorschools' heeft binnen verschillende schoolsystemen een verschillende invulling. Maar aangezien de gereviewde literatuur vrijwel uitsluitend de Verenigde Staten betreft, gaat het om kinderen in de leeftijd vanaf nul tot en met vier jaar. Vanaf vijf jaar bezoeken de meeste Amerikaanse kinderen een 'Kindergarten', een onderwijssoort die in enkele opzichten vergelijkbaar is met

Nederlandse kleuterklassen.

Aan de ontwikkeling van kinderen zijn vanzelfsprekend heel diverse facetten te onderscheiden. In de onderwijskundige literatuur wordt veelal een driedeling gemaakt: het cognitieve domein (waaronder intelligentie, ontluikende geletterdheid en ontluikende gecijferdheid), het sociaal-emotionele domein en het motorische domein (Seifert & Hoffnung, 1994). Wij hebben binnen alle drie domeinen naar relevante reviews gezocht, maar de aandacht is, zoals bekend, ongelijk verdeeld. De meeste aandacht is in het verleden uitgegaan naar het cognitieve domein, op grote afstand gevolgd door de beide andere domeinen.

De laatste specificatie betreft de kinderen voor wie de programma's bedoeld zijn. In eerste instantie wordt deze leeftijdsgroep als ongedeelde populatie beschouwd. Maar binnen deze populatie is een subpopulatie die van onderwijskundigen bijzondere aandacht gekregen heeft: de kinderen die, in Amerikaanse termen gesproken, om economische, culturele of sociale redenen in een risicosituatie verkeren. Deze subpopulatie komt in belangrijke opzichten overeen met de doelgroep van het Nederlandse onderwijsvoorrangsbeleid. In feite zijn veel van de ontwikkelde stimuleringsprogramma's op deze subpopulatie gericht. Uitgesloten hebben wij een andere subpopulatie: kinderen die om medische of paramedische redenen een risicogroep vormen, waaronder doven, blinden/slechtzienden, te vroeg geboren kinderen, mentaal geretardeerden, kinderen met spraak/taalstoornissen, kinderen van drugsverslaafden, enzovoorts.

Daarmee komen we toe aan de *secundaire* vraag. In de meeste leerboeken wordt onderzoekers aangeraden de wetenschappelijke kwaliteit van hun bronnen grondig te analyseren, voordat men conclusies hieruit overneemt. Die aanbeveling geldt niet alleen voor primaire bronnen zoals onderzoeksrapporten en artikelen over afzonderlijke studies, maar ook voor secundaire bronnen zoals reviews. Voor secundaire bronnen is dit evenwel vaak moeilijk te doen. Het ontbreekt menig onderzoeker aan tijd om terug te gaan naar de oorspronkelijke bronnen, en, het vervolgens combineren van de gegevens uit soms tientallen onderzoeken is een heel bewerkelijke zaak. Omdat er in veel gevallen bovendien maar één review beschik-

baar is, heeft een individuele onderzoeker meestal geen andere keus dan zich geheel en al te verlaten op de conclusies die de reviewer aan het einde van zijn studie getrokken heeft. In de praktijk ziet men dan ook vaak dat conclusies uit reviews letterlijk en zonder kritische kanttekeningen worden overgenomen.

Dit is een risicovolle werkwijze. In de enkele gevallen dat reviews kritisch onder de loep zijn genomen, zijn de uitkomsten desastreuus. Zo laat Slavin (1984) zich uiterst kritisch uit over een achttal belangrijke meta-analyses uit de eerste helft van de jaren tachtig: "... every one I examined in detail contained sources of bias serious enough, in my judgment, to call into question one or more major conclusions" (p. 8). Een ander voorbeeld vormen White, Taylor en Moss (1992) die de vloer aanvegen met zes reviews naar het effect van 'parent involvement' in voorschoolse stimuleringsprogramma's. Ten aanzien van de veronderstelling dat het betrekken van ouders de effectiviteit van de programma's gunstig beïnvloedt, concluderen zij: "The data cited by prominent reviewers of the early intervention research literature are largely contradictory or irrelevant to such a hypothesis" (p. 118).

In ons geval verkeerden we in de gelukkige omstandigheid dat er over één terrein verscheidene min of meer recente reviews beschikbaar waren. Dat biedt de gelegenheid ze naar hun aanpak en resultaten te vergelijken. De secundaire onderzoeksvraag luidde: 'Hoe degelijk zijn de reviews - ten aanzien van het in de primaire vraagstelling aangeduide probleemgebied - uitgevoerd?' Bij de beantwoording van deze vraag is het onmogelijk gebleken terug te gaan naar de oorspronkelijke of primaire studies. De werklust die hiermee gepaard zou gaan, lag ver boven onze mogelijkheden. Bij de beoordeling van de reviews hebben we alleen gebruik gemaakt van gegevens die in de reviews zelf vermeld zijn. De wijze waarop we te werk gegaan zijn, wordt in paragraaf 2.3 nader uitgewerkt.

## 2.2 Selectie van reviews

Wij hebben een poging gedaan alle relevante reviews van na 1980 bij onze review te betrekken. Reviews van vóór het aangegeven jaar zijn in zekere zin te beschouwen als behorende tot een eerdere generatie. Hun conclusies zijn

overgenomen, en, indien nodig, gecorrigeerd in latere overzichten. Dit geldt bijvoorbeeld voor de eerste review van Head Start projecten (Westinghouse Learning Corporation, 1969), die inmiddels is achterhaald door de review van McKey, Condelli, Ganson, Barrett, McConkey en Plantz (1985).

Om de reviews op het spoor te komen, zijn searches uitgevoerd in de bestanden van ERIC en PsycLit. Tevens zijn jaargangen van de relevante tijdschriften van de 'American Educational Research Association' doorgenomen, evenals handboeken en jaarboeken, zoals de 'Annual Review of Psychology' en de 'Review of Research in Education'. Onze speurtocht is beperkt gebleven tot de Engelstalige literatuur. Reviews die als prepublicatie zijn uitgebracht (bijvoorbeeld Harrell, 1983), zijn buiten beschouwing gebleven. Evenmin is Lazar en Darlington (1982) opgenomen, omdat het 'Consortium for Longitudinal Studies' (1983) nieuwe data bevat - namelijk op basis van een extra follow-up in 1980 - die in Lazar en Darlington (1982) niet beschreven worden.

Ten aanzien van de reviews met een bredere vraagstelling dan wij hier hanteren, is als volgt gehandeld. Indien in zulke reviews het voor ons relevante gedeelte onderscheidbaar was van het gedeelte dat onze vraagstelling te buiten ging, hebben wij ze toch opgenomen. Dit geldt bijvoorbeeld voor White (1985-1986) en Faran (1990) die zowel programma's voor achterstandskinderen als voor gehandicapte kinderen bespreken. De niet van toepassing zijnde delen zijn buiten beschouwing gebleven. Evenwel, in de review van Clark (1988), waarin alle Engelse onderzoek naar voorschoolse stimulering opgenomen is, bleek het onmogelijk een gedeelte af te zonderen dat betrekking had op de effectiviteit van programma's. Deze review is daarom door ons niet opgenomen.

Uiteindelijk zijn tien reviews uitgekozen. In de literatuurlijst zijn deze met een asterisk gemarkeerd. Naar ons beste weten hebben we daarmee alle reviews die aan de gestelde criteria voldeden bij onze analyse betrokken.

## 2.3 Verdere werkwijze

Bij de beoordeling van de reviews zijn vier hoofdcriteria onderscheiden: selectie, codering, integratie en conclusies. Deze criteria zijn ontleend aan door Kulik en Kulik (1989) op-

gestelde richtlijnen. Per hoofdcriterium wordt een korte toelichting gegeven.

### *Selectie*

Een goede review voldoet aan een aantal eisen wat betreft de selectie van de studies die als primaire bronnen worden gebruikt. Hoofdpunt is dat de reviewer expliciete criteria hanteert om studies wel of niet op te nemen. Deze criteria moeten afgeleid zijn uit de centrale vraag van de review. De criteria zijn als het ware de meest concrete operationalisatie van de onderzoeksvraag, waarin bijvoorbeeld de afhankelijke en onafhankelijke variabelen nader worden gespecificeerd. Aanvullend kunnen criteria geformuleerd worden over de methodologische kwaliteit van de studies, maar het nut hiervan is omstreden. Vanzelfsprekend behoren de gehanteerde criteria vermeld te worden en behoort de reviewer er zich aan te houden. Binnen deze criteria dient naar volledigheid gestreefd te worden.

### *Codering*

In een goede review worden relevante kenmerken van de studies gecodeerd. Daardoor is het mogelijk de relaties tussen kenmerken van de studies en de uitkomsten statistisch te analyseren. Relevante kenmerken betreffen de aard van de afhankelijke variabelen en de gehanteerde operationalisaties, de aard en de lengte van de experimentele behandeling, de resultaten van de studie en een aantal aspecten van de experimentele opzet (bijvoorbeeld het type opzet). Daarnaast kunnen variabelen die de publicatiestatus betreffen (bijvoorbeeld het jaar van publicatie, het type publicatie), worden gecodeerd. Het spreekt voor zich dat de reviewer aandacht moet besteden aan de betrouwbaarheid van de coderingsprocedures.

### *Integratie*

Bij de integratie van de studie-uitkomsten zullen veelal één of meer maten voor de sterkte van de gevonden relaties een rol spelen. Dit kan een correlatie-coëfficiënt zijn, of een op één of andere wijze gestandaardiseerde maat voor de effectgrootte. Bij het combineren van de gekozen maten moeten de juiste correcties worden toegepast. Voorts moet verhinderd worden dat studies met veel afhankelijke variabelen een te groot gewicht krijgen. Eenheid van analyse

dient de studie te zijn, niet de effectgrootte. In voorkomende gevallen dienen daarom in een vóórfase de uitkomsten per studie 'gepooled' te worden.

### *Conclusies*

Een laatste eis is dat reviewers geen conclusies trekken die niet door hun analyses worden ondersteund. De conclusies dienen door de beschreven uitkomsten ondersteund te worden.

## 3 Resultaten

### 3.1 De eerste oriëntatie

Tabel 1 biedt een eerste oriëntatie op de tien reviews die onder de loep genomen zijn. Wat betreft het aantal primaire studies waarop de reviews zich baseren, zijn er grote verschillen. Stallings en Stipek (1986) bespreken - naast de review van het Consortium uit 1983 - slechts twee primaire studies. Het andere uiterste is McKey et al. (1985) die 210 studies bespreken. Dat lijkt een gigantisch aantal. Maar van deze 210 studies bleken er slechts 76 te zijn die voldoende gegevens bevatten voor een meta-analyse.

We hebben een poging gedaan de overlap tussen de reviews wat betreft de primaire bronnen te bepalen. Maar omdat niet alle reviews een volledige lijst van hun bronnen bevatten, is dit maar gedeeltelijk gelukt. Met name de reviews van White (1985-1986) en van White, Taylor en Moss (1992) missen een lijst met gereviewde bronnen. De review van het 'Consortium for Longitudinal Studies' (1983) is in zekere zin als een belangrijke kern te beschouwen. In alle latere reviews, waarschijnlijk ook in die van White (1985-1986) en van White, Taylor en Moss (1992), komen één of meer van de door het 'Consortium for Longitudinal Studies' gereviewde bronnen terug. Bovendien zijn er twee reviews die een bespreking wijden aan de review van het 'Consortium for Longitudinal Studies' zelf, namelijk Stallings en Stipek (1986) en Sylva en Wiltshire (1993). De review van McKey et al. (1985) blijkt tamelijk los te staan van de latere reviews. Voor zover wij konden nagaan, is van de hierin opgenomen bronstudies geen enkele in latere reviews terecht gekomen. De review zelf wordt alleen in Karweit (1989) aan de orde ge-

**Tabel 1***Beschrijvende kenmerken van tien reviews*

Reviews	Aantal studies	Overlap met eerdere reviews <sup>b</sup>	Leerlingenpopulatie	Interventieperiode
'Consortium for Longitudinal Studies' (1983)	11	niet van toepassing	Amerikaans, achterstands-groeperingen, 0-5 jr.	tussen 1961 en 1969
Clarke-Stewart & Fein (1983)	ca. 80 <sup>a,c</sup>	8 maakten deel uit van 'Consortium for Longitudinal Studies' (1983)	Niet vermeld, vermoedelijk voornamelijk Amerikaans, gemengd qua SES status, 0-5 jr.	tussen 1955 en 1980
McKey et al. (1985)	210	2 studies maakten deel uit van 'Consortium for Longitudinal Studies' (1983)	Amerikaans, achterstands-groeperingen, 3-4 jr.	tussen 1965 en 1980
White (1985-1986)	99 <sup>a</sup>	onbekend (geen lit. lijst beschikbaar)	Amerikaans, achterstands-groeperingen, ?-5 jr.	merendeels tussen 1960 en 1982
Stallings & Stipek (1986)	3 (w.o. één review) <sup>a</sup>	volledige overlap met 'Consortium for Longitudinal Studies' (1983)	Amerikaans, achterstands-groeperingen, 0-5 jr.	tussen 1961 en 1964
Goldring & Presbey (1986)	8	volledige overlap met 'Consortium for Longitudinal Studies' (1983)	Amerikaans, achterstands-groeperingen, 0-5 jr.	tussen 1961 en 1969
Karweit (1989)	6 (w.o. één review) <sup>a</sup>	de review is McKey et al.; vier van de vijf studies maakten deel uit van 'Consortium for Longitudinal Studies' (1983)	Amerikaans, achterstands-groeperingen, 3-4 jr.	tussen 1961 en 1980
Farran (1990)	24 <sup>a</sup>	7 studies maakten deel uit van 'Consortium for Longitudinal Studies' (1983)	voornamelijk Amerikaanse achterstandsgroeperingen, 0-5 jr.	tussen 1961 en 1984
White, Taylor & Moss (1992)	97 <sup>a</sup>	onbekend (geen lit. lijst beschikbaar)	Amerikaans, achterstands-groeperingen, 7-5 jr.	tussen 1961 en 1982
Sylva & Wiltshire (1993)	17 (w.o. drie reviews)	1 studie maakte deel uit van 'Consortium for Longitudinal Studies' (1983)	deels Amerikaans, deels Engels, deels achterstands-groeperingen, deels dwarsdoorsnede uit bevolking	tussen 1961 en 1990

<sup>a</sup> Vermeld wordt het aantal studies binnen het voor ons relevante deel.

<sup>b</sup> Bij dit kenmerk wordt de review van het 'Consortium for Longitudinal Studies' (1983) als de eerste beschouwd.

<sup>c</sup> Het aantal studies is niet precies te bepalen.

steld. Ten slotte, alleen Farran (1990) en Sylva en Wiltshire (1993) bevatten een beduidend aantal nieuwe - niet in het 'Consortium for Longitudinal Studies' (1983) of in McKey et al. (1985) besproken - bronnen.

Ook al zijn de reviews zelf van redelijk recente datum, de voorschoolse stimuleringsprogramma's waarover gerapporteerd wordt zijn beduidend ouder. De programma's die besproken worden in de review van het 'Consortium

for Longitudinal Studies' (1983) zijn uit de jaren zestig, die van McKey et al. (1985) uit 1965 tot aan 1980. Alleen Sylva en Wiltshire (1993) bespreken programma's die minder dan tien jaar geleden zijn uitgevoerd.

Kenmerkend is het vrijwel exclusieve accent op programma's die in de Verenigde Staten zijn ontwikkeld voor kinderen in achterstandssituaties. Gezien de heersende Amerikaanse opvattingen gaat het om kinderen uit

Tabel 2

Beoordelingen van tien reviews (+ = positief oordeel; - = negatief oordeel of geen oordeel mogelijk)

Reviews	Criteria			
	Selectie	Codering	Integratie	Conclusies
'Consortium for Longitudinal Studies' (1983)	+	+	+	-
Clarke-Stewart & Fein (1983)	-	-	-	-
McKey et al. (1985)	+	+	+	+
White (1985-1986)	-	-	-	+
Stallings & Stipek (1986)	-	-	-	-
Goldring & Presbey (1986)	+	-	-	-
Karweit (1989)	-	-	-	-
Farran (1990)	-	-	-	-
White, Taylor & Moss (1992)	-	-	-	+
Sylva & Wiltshire (1993)	-	-	-	-

arme gezinnen, voornamelijk als 'black' aangeguid. Clarke-Stewart en Fein (1983) en Sylva en Wiltshire (1993) zijn de enige twee reviews die niet exclusief focussen op programma's voor kinderen in achterstandssituaties. Sylva en Wiltshire bespreken enkele programma's voor een dwarsdoorsnede uit de Engelse bevolking, maar maken in hun review niet duidelijk of de aard van de populatie waar de interventieprogramma's voor bedoeld zijn, van invloed is op de effectiviteit ervan.

### 3.2 De reviews afzonderlijk besproken

In deze paragraaf worden de afzonderlijke reviews kort besproken. Een meer uitgebreide bespreking is nu eenmaal niet mogelijk binnen de ruimte die een tijdschrift kan bieden. In de besprekingen wordt verwezen naar beoordelingen ten aanzien van de vier in paragraaf 2.3 onderscheiden criteria voor kwaliteit. De oordelen zijn in Tabel 2 samengevat.

'Consortium for Longitudinal Studies' (1983). Halverwege de jaren zeventig werd een consortium gevormd door onderzoekers die de lange-termijneffecten van hun vroege-interventieprogramma's - uitgevoerd in de jaren zestig - wilden vaststellen. In het consortium werden elf (van de in totaal 15 geïdentificeerde) afzonderlijke studies ingebracht. Overeengekomen werd zowel in 1976 als in 1980 follow-up data te verzamelen. Ondanks alle in het werk gestelde pogingen zoveel mogelijk van de oorspronkelijke proefpersonen te bereiken, bedroeg het aantal proefpersonen in 1980 nog slechts ongeveer 1100, dat wil zeggen minder dan één derde van het aantal oorspronkelijke proefpersonen. Bij de laatste follow-up varieerden de proefpersonen qua leeftijd van

10 tot meer dan 20 jaar. Ook inhoudelijk varieerden de programma's aanzienlijk (onder andere wat betreft de leeftijd waarop de interventie destijds plaats vond, de betrokkenheid van de ouders en de setting). In tegenstelling tot de andere hier te bespreken reviews zijn in deze review de oorspronkelijke data geanalyseerd. Het mediane effect, onmiddellijk na de interventie, werd geschat op 7 IQ-punten (gebaseerd op 7 van de 11 projecten), afnemend tot 3 IQ-punten drie tot vier jaar later. Op schoolvorderingen werden nauwelijks verschillen aangetoond. Opvallend is dat er op schoolloopbaanvariabelen wel verschillen aangetoond werden, en ook dat deze verschillen in de loop der jaren steeds groter werden (minder verwijzingen naar speciaal onderwijs, minder zittenblijven). Pogingen om de variatie in de afzonderlijke projectuitkomsten te verklaren, liepen op niets uit: "All of these programs were effective, and none were more or less effective than others as measured by our school outcome indicators" (p. 462). Derhalve luidde de conclusie dat vroege interventies, ongeacht de vormgeving van het interventieprogramma, een gunstig effect hebben op het IQ (alleen een onmiddellijk effect) en op schoolloopbaanvariabelen (verwijzingen, zittenblijven). In deze conclusie komt onvoldoende tot uiting dat de gegevens en de analyses een tamelijk gefragmenteerde indruk maken. De terugloop in proefpersonen is aanzienlijk, de meeste vergelijkingen zijn gebaseerd op slechts een deel van de studies, de studies verschillen onderling aanzienlijk in de aard van hun onafhankelijke variabele en in evaluatie-opzet. Wij hebben deze review daarom op het vierde criterium (conclusies) als onvoldoende beoordeeld.

Het hoofdstuk van *Clarke-Stewart en Fein (1983)* in het 'Handbook of Child Psychology' biedt een breed perspectief op voorschoolse stimuleringsprogramma's. De review van onderzoek naar de effecten van zulke programma's bestrijkt slechts een gedeelte van het hoofdstuk. In dit gedeelte worden de effecten onderverdeeld naar vier domeinen. Wij bespreken hier alleen het effectonderzoek binnen het sociale en het intellectuele domein. Aan dit deel ligt een grote hoeveelheid onderzoek ten grondslag. Binnen het sociale domein worden circa 45 verschillende literatuurverwijzingen gegeven, binnen het intellectuele domein gaat het om ongeveer 35 verwijzingen. Het aantal onderliggende studies is niet precies te bepalen, omdat het onduidelijk is of elke verwijzing correspondeert met een afzonderlijke effectstudie. De auteurs verstrekken geen toelichting op de criteria die aan de keuze van de gereviewde studies ten grondslag hebben gelegen. De review is kwalitatief van aard. Dat wil zeggen, bij het coderen van de afzonderlijke studies en de integratie zijn geen objectieve of kwantitatieve procedures gevolgd. Slechts incidenteel wordt melding gemaakt van de significantie of van de grootte van de gevonden effecten. Voor het sociale domein luidt de hoofdconclusie dat kinderen na afloop van een programma sociaal competent of rijper zijn. Onduidelijk is of deze conclusie een referentiegroep veronderstelt van kinderen die géén programma gevolgd hebben. Voor het intellectuele domein wordt geconcludeerd dat de programma's een consistent positief effect laten zien dat meestal kleiner is dan één standaarddeviatie. Doordat de auteurs zo weinig inzicht geven in de aard van het onderliggende onderzoek en in de door hen gevolgde integratieprocedure, valt de validiteit van deze conclusies niet na te gaan. Wij hebben deze review op alle vier de criteria als onvoldoende gekwalificeerd.

*McKey et al. (1985)* is een review van al het tussen 1967 en 1982 verrichte onderzoek - gepubliceerd en ongepubliceerd - naar effecten van Head Start programma's. De aandacht ging uit naar effecten op kinderen (wat betreft hun cognitieve en sociaal-emotionele ontwikkeling en hun gezondheid), op gezinnen en op de lokale gemeenschappen waarbinnen de programma's geïmplementeerd werden. Wij be-

spreken hier alleen het deel dat betrekking heeft op de effecten op de kinderen zelf. De speurtocht naar relevant onderzoek is uitputtend geweest. De gehanteerde selectie-criteria en de coderingsprocedures worden uitgebreid verantwoord en maken een degelijke indruk. Van de 210 opgespoorde studies bleken er slechts 76 voldoende gegevens te bevatten voor een meta-analyse. Ter verklaring van de variatie in de gerapporteerde effectgroottes werd een uitgebreide set predictoren in stelling gebracht, waaronder het type onderzoeksoptzet ('pre-post' versus 'control-treatment'), kenmerken van de programma's en van de deelnemende kinderen. De conclusie is dat kinderen door Head Start een onmiddellijk voordeel ondervinden op alle drie de typen effectmaten (cognitie, sociaal-emotionele ontwikkeling, gezondheid). De gemiddelde effectgrootte binnen het cognitieve domein bedroeg .5 sigma, en voor het sociaal-emotionele domein .2 à .3 sigma. Evenwel, de effecten werden van jaar tot jaar kleiner. Twee tot drie jaar na afloop van het programma, bleek de winst voor Head Start kinderen geheel verdwenen te zijn. Voor deze laatste conclusie konden de reviewers overigens nog op slechts minder dan tien primaire studies terugvallen. Op alle vier kwaliteitscriteria hebben wij deze review een positieve score gegeven.

*White (1985-1986)* rapporteert over een database met meer dan 300 verschillende studies naar de effectiviteit van vroege-interventieprogramma's. Hiervan hebben er 99 betrekking op programma's voor achterstandskinderen, merendeels uitgevoerd in de jaren zestig en zeventig. Onze bespreking regardeert alleen dit deel van de review. Hoewel White volledigheid suggereert, heeft hij beduidend minder studies gereviewed dan McKey et al. (1985), die alleen onderzoek naar Head Start programma's (noodzakelijkerwijs na 1965) geanalyseerd hebben. Dit gegeven roept twijfel op aangaande de geclaimde volledigheid. Elke studie is op circa 90 kenmerken gecodeerd (onder meer wat betreft de proefpersonen, de interventie en de methodologische kwaliteit van het onderzoek). Desondanks is de review opvallend weinig specifiek aangaande de afhankelijke en onafhankelijke variabelen in de primaire studies. De auteur schat het onmiddellijke effect van de interventie op ongeveer .5

sigma, in de loop van drie tot vijf jaar na de interventie afnemend tot nul. In deze schatting zijn effecten op zeer verschillende afhankelijke variabelen samengevoegd. Ondanks het grote aantal gecodeerde variabelen per studie slaagt hij er niet in de variantie in de gerapporteerde effectgroottes te verklaren. Aangaande het soort programma dat het meest effectief is, levert de review feitelijk weinig op. Wij hebben deze review alleen op het vierde criterium, aangaande de correspondentie tussen de gepresenteerde feiten en de getrokken conclusie, positief gewaardeerd. Aangetekend zij hierbij, dat de negatieve beoordeling voornamelijk is ingegeven door de onvolledigheid van de rapportage wat betreft de gehanteerde zoek-, codeer- en integratieprocedures. Niet uit te sluiten is dat deze procedures op zichzelf beschouwd adequaat zijn uitgevoerd.

*Stallings en Stipek (1986)* Deze review bestrijkt het hele lager onderwijs. Wij bespreken hier alleen het deel dat betrekking heeft op voorschoolse stimulering. De auteurs bespreken alleen longitudinaal onderzoek. Methodologische overwegingen hebben een rol gespeeld bij de keuze van hun bronstudies (zie p. 728), maar omdat ze hun criteria niet prijs geven, is de kwaliteit van hun selectieprocedure niet te controleren. Aan één van de Consortium-studies wordt uitgebreide aandacht besteed, omdat "... the Perry Preschool Program stands out in having significant long-term findings ..." (p. 729). Door deze aanpak wordt bewust bijgedragen aan het verschijnsel 'publication bias' (een eenzijdige keuze van de te reviewen bronnen). De review is kwalitatief van aard, waarbij de resultaten op een niet-statistische wijze worden geïntegreerd. De reviewers laten zich niet uit over de grootte van de gerapporteerde effecten. In het slotwoord worden allereerst de conclusies van het 'Consortium for Longitudinal Studies' (1983) met instemming aangehaald (p. 740). Er wordt aan toegevoegd dat activiteiten om het gezin bij school te betrekken een gunstig effect hebben op de slaagkans van kinderen. Deze laatste conclusie berust slechts op twee studies die een selecte keuze vormen uit een veel grotere groep studies naar het effect van ouderparticipatie. Een andere conclusie over het gunstige effect van Montessori-programma's op de cognitieve ontwikkeling berust op slechts één studie,

waarin niet duidelijk is geworden waarom dit effect wel voor vierjarige jongens maar niet voor vierjarige meisjes is gevonden. Deze review vertoont op alle vier criteria serieuze tekortkomingen.

*Goldring en Presbey (1986)*. Centrale vraag van deze review is hoe sterk het effect van preschool programma's is op vier met name genoemde afhankelijke variabelen: IQ, lees- en rekenprestatie ('in grade 4') en het percentage leerlingen met een vertraagde schoolloopbaan. Als bronstudies heeft men de acht Consortium studies gekozen die relevante data bevatten. De effectgroottes van deze studies worden afzonderlijk gerapporteerd. Voor bijvoorbeeld IQ variëren ze van -.19 tot en met .86. Deze variatie is aanzienlijk, maar de auteurs concluderen desalniettemin dat de studies een homogeen positief effect laten zien (algemene effectgrootte voor IQ, vier tot zes jaar na de interventie: .42). Zij gaan voorbij aan grote verschillen in de aard van de voorschoolse stimuleringsprogramma's en ze doen geen pogingen de variatie in effectgrootte te verklaren. Het weglekken van de effecten onder invloed van de tijd wordt door de auteurs met statistische argumenten ontkend, maar hun figuren laten tegelijkertijd zien dat de 95-procents betrouwbaarheidsintervallen de nul insluiten. Hun conclusies stemmen niet overeen met de gepresenteerde feiten. Deze review schiet tekort wat betreft drie van de vier criteria.

*Karweit (1989)* bespreekt de korte- en lange-termijneffecten van voorschoolse stimuleringsprogramma's. Wat betreft volledigheid legt de auteur een sterke claim: "This chapter reviews the major literature on the effects of preschool." (p. 76). Maar in feite bespreekt ze slechts vijf primaire bronnen (waarvan er vier ook door het 'Consortium for Longitudinal Studies' besproken zijn, namelijk Gray's 'Early training project', het 'Perry Preschool Project' van Schweinhart en Weikart, het Louisville experiment van Miller en Bizzel en het onderzoek van Karnes et al.) en daarnaast de meta-analyse van McKey et al. (1985). In de review staan twee soorten vergelijkingen centraal. Eerst wordt - op basis van drie studies - het effect van wel of geen 'preschool' programma besproken. Van twee van de drie studies worden effectgroottes vermeld, van de derde alleen de richting van de verschillen.



Karweits conclusie is dat voorschoolse stimulering tot een onmiddellijk positief effect leidt ter grootte van ongeveer 1.00 sigma, afnemend tot ongeveer .20 sigma aan het eind van 'grade 3'. Deze conclusie is niet in overeenstemming met de vermelde gegevens. De genoemde effectgrootte van 1.00 is in feite een weergave van het grootste gerapporteerde effect, en derhalve niet representatief voor de door Karweit aangedragen effectgroottes. In het tweede deel van de review worden de effecten van programma's onderling vergeleken. Daarbij komen geen consistente verschillen in effectiviteit aan het licht. De analyse van de verschillen tussen de programma's is erg oppervlakkig. Er is geen poging gedaan de verschillen op vaste of systematische punten te beschrijven. Karweits conclusie luidt dat de programma's elkaar in effectiviteit weinig ontlopen en dat ze allemaal de moeite waard lijken. Door onduidelijkheden in de vermelde gegevens is de validiteit van deze conclusie niet na te gaan. Wij zijn van oordeel dat haar review op alle vier criteria tekort schiet.

De review van *Farran (1990)* bestaat uit twee delen, een voor 'disadvantaged' en een voor 'handicapped children'. Wij bespreken alleen het eerste deel, waarin 24 bronstudies worden samengevat. Omdat bij sommige studies meer interventieprogramma's betrokken waren, is het aantal gereviewde programma's iets groter, namelijk 32. De auteur laat zich niet uit over de totstandkoming van zijn selectie. Daarom is niet na te gaan waarom er bijvoorbeeld van de elf oorspronkelijke Consortiumstudies door Farran slechts zeven besproken worden. De integratie van de uitkomsten is kwalitatief van aard, waardoor geen inzicht gegeven wordt in de grootte van de gerealiseerde effecten, noch in de kenmerken van de interventies of van de studies die de effectgroottes zouden kunnen beïnvloeden. De auteur concludeert dat zorgvuldig geplande en uitgevoerde vroege-interventieprogramma's een reeks gunstige effecten kunnen hebben ('a range of potential benefits'), maar dat interventies gecontinueerd moeten worden om uitdoving te voorkomen. In feite, zo vervolgt Farran, hebben achterstandsleerlingen meer aan een verbetering van het onderwijs in de eerste vier jaren van de basisschool dan aan vroege-interventieprogramma's. De door hem bespro-

ken studies leveren geen bevindingen op die voor deze slotconclusie relevant zijn. Naar ons oordeel voldoet Farrans review aan geen van de vier criteria.

*White, Taylor en Moss (1992)* reviewen een veel breder probleemgebied dan wij hier behandelen, namelijk de vraag of ouderparticipatie het effect van vroege-interventieprogramma's gunstig beïnvloedt. In een apart onderdeel, het enige deel dat wij hier aan de orde stellen, beantwoorden ze de vraag of interventieprogramma's mét een ouder-component sterkere effecten teweegbrengen dan programma's zónder zo'n component. Uit een in het begin van de jaren tachtig aangelegd bestand van onderzoeken (met ruim 350 studies) werden alle studies geselecteerd die voor deze vraag relevant materiaal bevatten. Vermoedelijk betreft het hier dezelfde studies die ook door de eerder besproken White (1985-1986) gereviewed zijn. Elke studie is op vele tientallen aspecten gecodeerd, onder meer wat betreft de afhankelijke en onafhankelijke variabelen en de kwaliteit van de onderzoeksopzet. Als uitkomsten van de studies werden effectgroottes bekeken. Voor details over de gehanteerde procedures wordt naar tussentijdse rapporten verwezen. Wij zijn er niet in geslaagd deze rapporten te bemachtigen, zodat we slechts kunnen constateren dat de beschikbare rapportage onvoldoende gegevens bevat om de kwaliteit van de coderingen en van de gegevensintegratie te beoordelen. Beperken we ons tot de effectgroottes voor studies met achterstandskinderen waarbij een 'center-based' aanpak werd gehanteerd, dan bedroegen deze gemiddeld .52, respectievelijk .53 voor de studies met, respectievelijk zonder oudercomponent. Deze cijfers hebben alleen betrekking op de 43 studies met een hoge interne validiteit. De 54 studies met een lage interne validiteit lieten wél een verschil zien. De conclusie van de auteurs luidt dat programma's mét een oudercomponent géén sterkere effecten laten zien dan de programma's die zo'n component missen. Deze conclusie blijft dicht bij de feiten zoals in het artikel weergegeven. Gezien de overeenkomsten met de review van White (1985-1986) hebben wij dezelfde oordelen gegeven.

*Sylva en Wiltshire (1993)*. Centrale vraag in deze review is of voorschoolse stimulerings-

programma's een blijvend effect hebben op de latere levensloop van kinderen ('a lasting effect on the course of children's later lives'). De auteurs geven aan alleen de 'key studies' te bespreken, maar weiden over hun selectie verder niet uit. Geconstateerd moet worden dat slechts één van de besproken studies bij hun vraagstelling past (namelijk het Perry Preschool Project). Het merendeel van de studies evalueert alleen effecten op korte termijn en binnen het schoolse domein. De bespreking van de afzonderlijke studies is onsystematisch en niet kwantitatief. In sommige gevallen worden effectgroottes vermeld, in andere gevallen niet. De auteurs concluderen dat voorschoolse interventie op de korte termijn tot meetbare effecten leidt binnen het schoolse en sociale domein. Voorts, als de interventie van hoge kwaliteit is, zijn de effecten op het cognitieve en sociale vlak van blijvende aard. De effecten gelden voor kinderen uit alle sociale groepen, maar zijn het sterkst voor achterstandsleerlingen. Geen van deze conclusies wordt gestaafd door het besproken onderzoek. Hetzelfde geldt voor de andere door Sylva en Wiltshire getrokken conclusies betreffende onder meer het economische voordeel van 'high quality early education'. Op alle vier criteria waarderen wij deze review negatief.

#### 4 Conclusies en discussie

Wij beginnen de afsluitende paragraaf met de conclusie ten aanzien van de secundaire vraagstelling, de vraag naar de kwaliteit van de besproken reviews. Onze conclusie is dat de kwaliteit van de reviews over het geheel genomen teleurstellend laag is. Slechts twee van de tien reviews, die van het 'Consortium for Longitudinal Studies' (1983) en van McKey et al. (1985), hebben wij op drie of vier van onze criteria van een positief oordeel kunnen voorzien. De resterende acht reviews zijn op nagenoeg alle criteria als onvoldoende gekwalificeerd. Wij vinden dit een onthutsende conclusie, mede omdat de bekritiseerde reviews in gerenommeerde tijdschriften of handboeken verschenen zijn. In het discussiedeel van deze paragraaf komen we op deze kwestie nog terug.

Bij onze conclusie ten aanzien van de primaire vraagstelling betrekken we slechts de re-

views die de toets van onze kritiek kunnen doorstaan. Dat zijn er, zoals gezegd, maar twee. Wat valt er ten aanzien van de effectiviteit van voorschoolse stimuleringsprogramma's voor lering te trekken? We onderscheiden vijf deelconclusies.

Ten eerste, direct na afloop van de interventie laten IQ-testen een gemiddelde winst van ongeveer .5 sigma zien, zijnde ongeveer 7 IQ-punten. Deze winst verdwijnt betrekkelijk snel ('Consortium for Longitudinal Studies', 1983: na het eerste jaar is de winst ongeveer gehalveerd; McKey et al., 1985: na één jaar is de winst nagenoeg helemaal verdwenen).

Ten tweede, wat betreft de vorderingen binnen het schoolse domein (schoolrijpheid, lezen, rekenen) blijkt alleen uit de review van McKey et al. (1985) een onmiddellijke winst van ongeveer .5 sigma. Deze winst ebt in ongeveer twee jaar weg. De review van het 'Consortium for Longitudinal Studies' (1983) laat zien dat een onmiddellijke winst op dit vlak soms wel, maar soms ook niet aantoonbaar is. De overheersende tendens is dat een eventueel effect op dit domein zich hoogstens tot en met 'grade 1' (groep 3 in ons basisonderwijs) uitstrekt.

Ten derde, gunstige effecten op de schoolloopbaan van de leerlingen (wat betreft verwijzingen naar speciaal onderwijs en zittenblijven) worden in slechts enkele studies aangetoond. Dergelijke effecten zijn derhalve omstreden, mede omdat het niet plausibel is dat leerlingen een gunstiger schoolloopbaan zouden weten te realiseren, zonder de daarbij behorende winst op het vlak van de schoolvorderingen.

Ten vierde, het onmiddellijke effect binnen het sociaal-emotionele domein varieert volgens McKey et al. (1985) van .4 sigma (namelijk voor sociaal gedrag) tot .2 sigma (namelijk voor zelfvertrouwen en prestatie-motivatie). De winst verdwijnt na het tweede jaar. Het 'Consortium for Longitudinal Studies' (1983) heeft binnen dit domein geen metingen gerapporteerd.

Ten vijfde, het is duidelijk dat de stimuleringsprogramma's ten aanzien van hun doelstellingen en werkwijze enorm van elkaar verschillen. Maar de consequenties van deze verschillen voor de effectiviteit van de programma's zijn nagenoeg onbekend. Het 'Con-

sortium for Longitudinal Studies' (1983) suggereert dat alle programma's even effectief zijn, maar deze conclusie wordt niet door de aangevoerde feiten gestaafd. De review van McKey et al. (1985) levert als enige houvast dat programma's met een sterk gestructureerde, schoolse gerichtheid effectiever zijn dan programma's die een dergelijke gerichtheid missen.

In het discussiegedeelte van deze paragraaf stellen we drie onderwerpen aan de orde.

Het eerste onderwerp is het belang van de aangetoonde effecten. Op zichzelf lijken effecten ter grootte van een halve sigma natuurlijk zeer de moeite waard. Maar de wetenschap dat de effecten in betrekkelijk korte tijd weglekken, zou aanleiding kunnen geven tot een geringschattende houding. Wij menen dat zo'n houding niet terecht is, zolang niet duidelijk is wat de achtergrond van dit weglekken is. Het is opvallend dat de besproken reviews hierover nauwelijks theoretiseren. Zigler, Styfco & Gilman (1993, p. 21) brengen naar voren dat de ontwikkeling van kinderen - ná de voorschoolse periode - vanzelfsprekend van veel meer grootheden afhankelijk is dan het éénjarige voorschoolse programma wat ze al dan niet gevolgd hebben. Invloeden van school en huis overschaduwden al snel de winst die kinderen als vierjarige ten gevolge van een voorschools programma geboekt hebben. Steeds meer wint het inzicht veld dat vroege interventie alleen niet genoeg is om tekorten in de thuisomgeving te verhelpen. Zigler (1987) verwoordt dit als volgt: "We simply cannot inoculate children in one year against the ravages of a life of deprivation" (p. 258). Nodig zijn vervolprogramma's die een ingebouwde compensatie bieden voor kinderen die thuis niet optimaal in hun ontwikkeling ondersteund worden. Longitudinaal onderzoek naar de effecten van voorschoolse stimuleringsprogramma's - zonder dat aan kinderen zulke vervolprogramma's worden aangeboden - vormt een miskenning van ontwikkelingspsychologische inzichten. Het weglekken van de winst pleit in feite meer voor het ontwerpen en evalueren van vervolprogramma's dan tot een geringschatting van het belang van voorschoolse stimulering.

De relevantie van deze vijf bevindingen

voor het Nederlandse onderwijsbeleid - het tweede discussieonderwerp - vormen feitelijk niet het onderwerp van dit artikel. Toch achten wij het raadzaam in dit verband een waarschuwend woord te laten horen. De hier beschreven bevindingen hebben betrekking op interventies, exclusief gericht op achterstandskinderen in de Verenigde Staten en voor het merendeel uitgevoerd tussen 1960 en 1980. Tussen 'toen' in de Verenigde Staten en 'nu' in Nederland zijn belangrijke verschillen die de generaliseerbaarheid ernstig kunnen beperken. We wijzen erop dat Nederlandse kleuters twee tot drie jaar onderwijs volgen, voordat ze in groep 3 met het meer formele taal- en rekenonderwijs starten. In de Verenigde Staten was - en is - de periode waarin kinderen geleidelijk in de eisen van het schoolstelsel kunnen ingroeien, veelal slechts één schooljaar, waarin bovendien veel kinderen slechts een halve-dagsprogramma volgen. Ook in andere opzichten zijn er verschillen. Het is onze indruk dat de kwaliteit van de thuisopvoeding van Nederlandse achterstandsleerlingen veel minder achterblijft ten opzichte van de vorming en educatie die leerlingen op school krijgen, dan van Amerikaanse achterstandskinderen. Als deze indruk juist zou zijn, zouden Nederlandse voorschoolse interventieprogramma's - zelfs als deze op de zwaarste achterstandscategorieën gericht zouden zijn - tot minder grote effecten leiden dan in de hier gereviewde literatuur gerapporteerd worden.

Rest ons de discussie van het derde onderwerp: hoe komt het dat hier de conclusie getrokken wordt dat zó veel van de besproken reviews in zo ernstige mate tekort schieten? Het antwoord zou kunnen zijn, dat wij 'unfaire' criteria hebben aangelegd. Zo zou men tegen kunnen werpen dat onze eisen te veel gemodelleerd zijn op een kwantitatieve reviewprocedure, ook wel aangeduid als meta-analyse ter onderscheiding van een traditionele review waarin de integratie voornamelijk kwalitatief van aard is. Het is inderdaad waar dat aan de hier gehanteerde criteria in feite alleen door meta-analyses voldaan kan worden. Maar bij deze tegenwerping passen twee argumenten. In de eerste plaats is de hoeveelheid te reviewen onderzoek zo groot, dat een voornamelijk kwalitatieve werkwijze omslachtig is en niet meer kan voldoen aan de reproduceerbaarheidseis.

In de tweede plaats is het gezien de publicatie-data van de reviews ook werkelijk 'fair' de meta-analytische criteria hier toe te passen. Wij wijzen erop dat alle reviews gepubliceerd zijn, nádat de meta-analytische werkwijze ontwikkeld is. De fundamenteën voor meta-analyse zijn immers al in de jaren zeventig gelegd door onder andere Glass (1976) en Rosenthal (1978). Wij houden derhalve staande dat onze criteria afgeleid zijn uit zeer algemene wetenschappelijke eisen aangaande objectiviteit en reproduceerbaarheid. Wij eisen, met andere woorden, dat denkbeeldige andere reviewers - gegeven eenzelfde vraagstelling en gegeven eenzelfde 'body of research' - tot dezelfde conclusies zouden moeten komen. Als aan deze eis niet tegemoet gekomen wordt, ontvalt de eigenlijke zin aan het wetenschapsbedrijf.

Kan het zijn dat reviewers op zich hun werk goed doen, maar dat redacties van tijdschriften en handboeken het hun onmogelijk maken om verantwoord verslag te doen? Men kan opmerken dat de beide 'goedgekeurde' reviews over veel rapportageruimte beschikken, terwijl de andere reviews zich moesten voegen binnen de beperkingen van een artikel in een tijdschrift of van een hoofdstuk in een handboek. Toch verwerpen wij voor de meeste van de afgekeurde reviews ook deze mogelijkheid. Slechts bij de reviews van White (1985-1986) en van White, Taylor en Moss (1992) hebben wij de mogelijkheid opengelaten, dat bestudering van onderliggende rapporten - waarover we in die gevallen niet konden beschikken - alsnog tot een positieve waardering zou kunnen leiden.

Aan de conclusie dat het merendeel van de reviews onbetrouwbaar is, valt ons inziens niet te ontkomen. De punten waarop de reviews tekort schieten, betreffen alle vier hier onderscheiden criteria.

Aangaande het criterium *selectie* treden twee problemen aan het licht. Opvallend is dat reviewers weinig aandacht besteden aan een zorgvuldige domeinafgrenzing vooraf, vóórdat het zoekproces wordt gestart. Die afgrenzing zou voort moeten komen uit een explicitering van de onderzoeksvraag. Diverse reviewers komen wat betreft hun vraagstelling niet verder dan een terloopse opmerking dat zij interventieprojecten reviewen. Specificaties die achterwege blijven, betreffen onder meer een type-ring van de aard van de afhankelijke

variabelen, de aard en de leeftijd van de leerlingenpopulaties waarop de interventies zich richten en het jaar waarin de interventies geëvalueerd zijn. Weliswaar bevatten de meeste reviews op deze punten een specificatie achteraf. Maar zo'n specificatie achteraf kan een specificatie vooraf niet vervangen, zo min als een beschrijving van het resultaat van een steekproeftrekking in de plaats kan treden van een specificatie van de populatie. Ook de gevolgde zoekprocedures - dat is het tweede probleem - blijven onderbelicht. Gevolg is dat de meeste reviewers geen garantie kunnen geven omtrent de volledigheid of representativiteit van het onderzoek waarop zij zich baseren.

Ten aanzien van het criterium *codering* is het opvallend dat in de meeste reviews geen kenmerken gecodeerd zijn, noch kenmerken betreffende de afhankelijke en onafhankelijke variabelen of de aard van de populatie, noch kenmerken die de publicatiestatus betreffen. Zo zijn de reviews in feite heel weinig informatief over de aard van de interventies waarop de onderzoeken betrekking hebben. Effecten van waarschijnlijk zeer uiteenlopende programma's worden zonder meer bij elkaar gevoegd. Toch is aannemelijk dat de vormgeving van een programma, met aspecten als de doelstellingen, de leertheoretische principes, de intensiteit en de duur, in sterke mate de effectiviteit bepalen. Wat we in feite constateren, is dat de reviews arm aan theorie zijn. Immers, theorie-gestuurde uitspraken over hoe jonge kinderen zich ontwikkelen dicteren idealiter welke variabelen er ten minste gecodeerd dienen te worden.

Wat betreft het criterium *integratie* is de grootste tekortkoming dat de uitkomsten van de studies hoofdzakelijk in kwalitatieve zin worden besproken. Effectgroottes en significantieniveaus spelen een geringe rol. Daardoor worden variaties in de uitkomsten van de afzonderlijke studies niet zichtbaar en vervalt de noodzaak voor deze variatie een adequate verklaring te postulieren.

Ook bij het laatste criterium, *conclusies*, zijn wij tot negatieve evaluaties gekomen. In een aantal reviews worden conclusies niet expliciet getrokken, maar alleen impliciet verwoord met weglating van relevante condities. Soms ook worden conclusies overgeslagen en stappen reviewers van de resultaten direct over

naar de implicaties voor de praktijk van de interventies. In meer dan één geval hebben wij geconstateerd dat conclusies niet in overeenstemming zijn met de gepresenteerde data. Het lijkt, kortom, of sommige reviewers subjectieve opvattingen als uitkomsten van een wetenschappelijk verantwoorde review presenteren.

## Noot

1. De opdracht is verstrekt door het Instituut voor Onderzoek van het Onderwijs (SVO), project 94807.

## Literatuur

(De besproken reviews zijn met een asterisk aangeduid.)

- Clark, M. M. (1988). *Children under five: educational research and evidence*. New York: Gordon and Breach Science Publishers.
- \*Clarke-Stewart, K. A., & Fein, G. G. (1983). Early childhood programs. In P. H. Mussen (Ed.), *Handbook of Child Psychology, formerly Carmichael's Manual of child psychology; Volume II: Infancy and developmental psychobiology* (pp. 917-1000). New York: Wiley.
- \*Consortium for Longitudinal Studies, The (1983). *As the twig is bent ...; Lasting effects of preschool programs*. Hillsdale, NJ: Lawrence Erlbaum.
- \*Farran, D. C. (1990). Effects of intervention with disadvantaged and disabled children: A decade review. In S. J. Meissels & J. P. Shonkoff, *Handbook of early childhood intervention* (pp. 501-539). Cambridge: Cambridge University Press.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- \*Goldring, E. B., & Presbey, L. S. (1986). Evaluating preschool programs: a meta-analytic approach. *Educational Evaluation and Policy Analysis*, 8, 179-188.
- Harrell, A. (1983). *The effects of Head Start on children's cognitive development: preliminary report* (Publication No. (OHDS) 83-31186). Washington, DC: DHHS Publications.
- \*Karweit, N. (1989). Effective preschool programs for students at risk. In R. E. Slavin, N. L. Karweit & N. A. Madden, *Effective programs for students at risk* (pp. 75-102). Needham Heights: Allyn & Bacon.
- Kulik, J. A., & Kulik, C.-L. C. (1989). Meta-analysis in education [special issue]. *International Journal of Educational Research*, 13(3).
- Lazar, I., & Darlington, R. (1982). The lasting effects of early education: a report from the Consortium for Longitudinal Studies. *Monographs of the Society for Research in Child Development*, 47 (nos. 2-3, serial no. 195).
- \*McKey, H. R., Condelli, L., Ganson, H., Barrett, B., McConkey, C., & Plantz, M. (1985). *The impact of Head Start on children, families and communities*. Washington, DC: DHHS publication.
- Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin*, 85, 185-193.
- Seifert, K. L., & Hoffnung, R. J. (1994). *Child and adolescent development* (third edition). Boston: Houghton Mifflin Company.
- Slavin, R. E. (1984, oktober). Meta-analysis in education: How has it been used? *Educational Researcher*, 13, 6-15.
- \*Stallings, J. A., & Stipek, D. (1986). Research on early childhood and elementary school teaching programs. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd edition) (pp. 727-753). New York: MacMillan.
- \*Sylva, K., & Wiltshire, J. (1993). The impact of early learning on children's later development: A review prepared for the RSA inquiry 'Start Right'. *European Early Childhood Education Research Journal*, 1, 17-40.
- Westinghouse Learning Corporation (1969). *The impact of Head Start: an evaluation of the effects of Head Start on children's cognitive and affective development*. ERIC (ED 036321).
- \*White, K. R. (1985-1986). Efficacy of early intervention. *The Journal of Special Education*, 19, 401-416.
- \*White, K. R., Taylor, M. J., & Moss, V. D. (1992). Does research support claims about the benefits of involving parents in early intervention programs? *Review of Educational Research*, 62, 91-125.
- Zigler, E. F. (1987). Formal schooling for four-year-olds? No. *American Psychologist*, 33, 254-260.
- Zigler, E., Styfco, S. J., & Gilman, E. (1993). The national Head Start program for disadvantaged preschoolers. In E. Zigler & S. J. Styfco (Eds.), *Head Start and beyond; a national plan for extended childhood intervention* (pp. 1-42). New Haven: Yale University Press.

## Auteurs

H. Blok en P.P.M. Leseman zijn verbonden aan het SCO-Kohnstamm Instituut van de Faculteit Pedagogische en Onderwijskundige Wetenschappen van de Universiteit van Amsterdam

Adres: Wibautstraat 4, 1091 GM Amsterdam

## Abstract

### **Effects of preschool programs: a review of reviews**

H. Blok, P.P.M. Leseman. *Pedagogische Studiën*, 1996, 73, 184-197.

This article deals with the question how preschool programs affect children's development. Ten reviews, published between 1983 and 1993, have been evaluated on four criteria: 1. the selection of the source studies; 2. the coding of the studies' characteristics; 3. the way the results were integrated and 4. the correspondence between the results and the conclusions. Eight out of these ten studies proved to be seriously defective with regard to two or more of the criteria. The two remaining reviews permit the following conclusions: a. IQ tests and achievement tests show an immediate gain of about half a sigma; b. the benefit on the social domain is about a quarter of a sigma; c. in most cases the benefit is no longer demonstrable after one to two years; d. only a few studies show an effect on the school career; e. little is known about the characteristics of the programs which affect the program's efficiency. Since most studies were carried out among American disadvantaged children between 1960 and 1980, it is difficult to generalize these conclusions to the present Dutch educational system.