Rapport

Deel 2

Effectieve interventies voor bestrijding onderwijsachterstanden

Wat werkt volgens de internationale literatuur en hoe is dit het best te onderzoeken?

Marco Algera, Heleen van der Stege (CED-Groep)
m.m.v. Marcel Spijkerman (SEOR) Kees Zandvliet (SEOR)
CED-Groep, Rotterdam
2013

Inhoud

Samenvatting 5

1 Achtergronden van het onderzoek 9

1.1 Inleiding 9

1.2 Onderzoeksvragen 10

1.3 Methode 10

1.4 Opbouw van het rapport 111

2 Onderzoek interventies vóór reguliere basisschooltijd: preschool en vve 13

2.1 Inleiding 13

2.2 Resultaten 14

2.3 Synthese 18

3 Onderzoek interventies tijdens reguliere basisschooltijden 21

3.1 Inleiding 21

3.2 Resultaten 21

3.3 Synthese 24

4 Onderzoek interventies na reguliere basisschooltijden: afterschool programs 27

4.1 Inleiding 27

4.2 Resultaten 27

4.3 Synthese 31

5 Onderzoeksdesigns 33

5.1 Inleiding 33

5.2 Resultaten 33

5.3 Overwegingen voor keuze onderzoeksdesign 366

5.4 Voorgestelde onderzoeksdesigns 41

5.5 Intermezzo: gebruik van reeds verzamelde data 42

5.5.1 Inleiding 422
5.5.2 PRIMA/COOL 43
5.5.3 DUO 46
5.5.4 Vve-monitor 48
5.5.5 Mogelijkheden voor effectmeting – een eerste verkenning 49
5.5.6 Haalbaarheid fase 1 van voorgesteld onderzoeksdesign 49
5.5.7 Conclusies mogelijkheden bestaande databestanden 49
5.6 Synthese: uitwerking van een onderzoeksdesign 50

6 Beschouwing 53

Bibliografie 56

Samenvatting

In Nederland staat het onderwijsachterstandenbeleid al jaren in de schijnwerpers. Dit niet alleen vanwege de omvang van de geïnvesteerde overheidsgelden, maar ook omdat het terugdringen van onderwijsachterstanden een groot maatschappelijk belang dient, om nog maar te zwijgen van persoonlijke belangen van leerlingen die het betreft.
Tal van interventies worden in binnen- en buitenland gepleegd om in het primair onderwijs onderwijsachterstanden bij leerlingen te voorkomen of te bestrijden. Het betreft kinderen die *at risk* zijn, vaak in samenhang met etniciteit en sociaaleconomische status. Het gaat niet om kinderen die vanwege lage intelligentie of een lichamelijke of psychische handicap niet kunnen meekomen in de klas. In dit rapport wordt beschreven van welke interventies door wetenschappelijk onderzoek de effectiviteit is aangetoond. De centrale vraagstelling van dit onderzoek luidt dan ook: *"Welke door scholen gepleegde interventies dragen bij aan een effectieve bestrijding van onderwijsachterstanden in het primair onderwijs?*". Speciale aandacht gaat hierbij uit naar hoe de effectiviteit te onderzoeken is. De hieruit afgeleide onderzoeksvragen worden hieronder beknopt beantwoord. Door een combinatie van zoektermen op te geven aan literatuurzoekmachines is getracht een omvattend beeld te krijgen van effectieve maatregelen in het primair onderwijs, zoals dat vooral geschetst wordt in *narrative* en systematische reviews vanaf het jaar 2000.

Welke interventies komen als effectief naar voren in internationale wetenschappelijke literatuur?
Om de veelheid aan informatie te structureren, is een indeling naar schoolperiode aangehouden. Er zijn interventies die *voorafgaand aan de basisschooltijd* uitgevoerd kunnen worden, ter voorkoming van onderwijsachterstanden in het primair onderwijs. Dit soort interventies is samen te brengen onder de noemer '*pre-schooling*'. In Nederland wordt gesproken van 'voor- en vroegschoolse educatie' (vve). Behalve de voorschoolse activiteiten voor peuters van 2,5 en 3 jaar op de peuterspeelzaal en in de kinderopvang, worden hiertoe ook programma's gerekend die worden aangeboden in groep 1 en 2 van de basisschool. De effectiviteit van bepaalde vormen van vve wordt in onderzoek aangetoond; hierin komt consensus over het nut van het investeren in de voor- en vroegschoolse periode naar voren. Enkele andere onderzoeken tonen echter geen effectiviteit van vve aan om onderwijsachterstanden te voorkomen. Het blijkt in grote mate van de implementatie en de kwaliteit van het gekozen programma af te hangen, van het opleidingsniveau van medewerkers en of het programma al dan niet intensief en langdurig is. Helaas wordt in veel onderzoek de exacte inhoud van een programma niet beschreven. Verder zijn buitenlandse programma's vaak niet te implementeren in de Nederlandse context (ander schoolsysteem, voor Nederlandse begrippen veel te intensief).
Een tweede categorie interventies betreft die welke *tijdens de reguliere basisschooltijden* verricht kunnen worden. Hier worden niet zo zeer specifieke interventies genoemd als wel voorwaarden voor effectiviteit. Het gaat dan onder andere om een sterke pedagogische visie, de aard van het schoolleiderschap, de schoolcultuur en ouderbetrokkenheid.
De derde soort interventies worden *na de reguliere schooltijden* ingezet: *afterschool* programma's. Ook dit type programma's kent een keur aan verschijningsvormen. Hierbij komen schakelklassen in beeld, waarin leerlingen met een achterstand (veelal op taalgebied, maar ook wel op rekenen) onder schooltijd bijgespijkerd worden. In Nederlands onderzoek blijken deze schakelklassen (in de vorm van een heel leerjaar extra) zich uit te betalen in hogere taal- en rekenscores, beter kunnen meekomen op de basisschool en een advies voor een hoger onderwijsniveau in het voortgezet onderwijs.
Juist omdat programma's, van welke van de drie groepen interventies dan ook, veelal zo contextspecifiek zijn, is het des te belangrijker te kijken naar de werkzame elementen daarvan. In dit rapport worden deze elementen beschreven vanuit de onderzoeksliteratuur en vervolgens in kaders samengevat. Deze kaders vertonen behoorlijk wat overlap met elkaar. Op pagina 7 worden de drie kaders in elkaar geschoven bij wijze van samenvatting. De letteraanduiding staat respectievelijk voor vve (V), onder schooltijd (S) en *afterschool* (A).

Op welke wijze werd de effectiviteit van interventies onderzocht en wat is de bewijskracht van deze onderzoeken?

Interventies werden op tal van manieren onderzocht. De zoektocht naar literatuur spitste zich om praktische redenen toe op reviews, omdat die tal van onderliggende onderzoeken samenvatten. Hierin worden allerlei soorten onderzoeksdesigns gebruikt. In systematische reviews worden onderzoeksresultaten statistisch samengevoegd, hetgeen eisen stelt aan zowel de rapportage als aan het onderzoeksontwerp, dat veelal sterker (met meer bewijskracht) is dan in *narrative* reviews wordt gebruikt. Sterke designs omvatten (ongenuanceerd gesteld) methoden, waarbij de interventie gecontroleerd en in verschillende 'gradaties' wordt toegediend aan een experimentele en een controlegroep, die aselect zijn samengesteld.

In *narrative* reviews worden onderzoeksresultaten voortvloeiend uit (ook) mindere sterke designs in verhalende zin beschreven. Hierbij kunnen ook kwalitatieve designs zitten, die doorgaans weinig zeggingskracht hebben qua representativiteit.

Uit beide typen reviews wordt lang niet altijd duidelijk welke designs aan de opgenomen onderzoeken ten grondslag lagen.

Welke valkuilen, voor- en nadelen zijn er bij het onderzoeken van de effectiviteit?

Uit de literatuurstudie blijkt dat er nogal wat voetangels en klemmen zijn bij effectiviteitsonderzoek. Veel auteurs concluderen dat ook. De gevonden valkuilen zijn onder andere:

- aard van onderzoek: het design van het onderzoek laat veelal niet toe dat er verantwoorde uitspraken over de causaliteit gedaan kunnen worden; het design is vaak correlationeel of kwalitatief van aard;
- onderzoekspopulatie: de steekproef is klein, de analyse-eenheid niet passend bij de onderzoeksvraag of bij het randomiseren;
- verstorende variabelen: hierop is te weinig zicht, hun invloed blijft ongemeten en dus ongecorrigeerd of er is sprake van besmetting van de controlegroep met de interventie;
- interventie; deze blijkt geregeld onvoldoende gestandaardiseerd of überhaupt maar voldoende beschreven in onderzoeksartikelen; dit terwijl de effectiviteit van de interventie staat of valt met de kwaliteit ervan; ook met de dosering van de interventie (intensief of niet) wordt vaak niet gevarieerd;
- uitkomstmaat: de afhankelijke variabele wordt vreemd genoeg lang niet altijd eenduidig gedefinieerd, soms wordt er te vroeg geëvalueerd, waardoor de interventie nog geen tijd heeft gehad zich in meetbare effecten te laten zien of deugt het meetinstrumentarium om verwachte effecten überhaupt te kunnen meten niet;
- setting: (buitenlandse) onderzoeksresultaten kunnen niet zonder meer geëxtrapoleerd worden naar bijvoorbeeld het Nederlandse onderwijssysteem.

De beste resultaten worden behaald door gebruik te maken van een longitudinaal, liefst gecombineerd onderzoeksdesign, van een mix van dataverzamelingsmethoden en afkomstig van verschillende groepen respondenten.

Hoe kan de effectiviteit van interventies worden onderzocht?

In dit rapport wordt uitvoerig ingegaan op de sterkte van bepaalde onderzoeksdesigns. Uiteraard vindt onderzoek niet in een vacuüm van een kantoor of laboratorium plaats, maar is het onderhevig aan contextuele factoren, waarvan geld en tijd doorgaans de meest beperkende zijn. Ook praktische en ethische overwegingen kunnen een beperkende rol spelen. In die zin moet gestreefd worden naar het maximaal haalbare. Een onderzoek met een methodologisch voortreffelijk design, maar dat onbetaalbaar en in de praktijk onuitvoerbaar is, en/of ethisch niet verantwoord is, blijft onuitgevoerd en levert dus geen resultaten op.

Er wordt in dit rapport dan ook een aanzet voor een onderzoeksopzet gegeven, dat hiermee rekening houdt. Hierbij wordt in een eerste fase gebruik gemaakt van landelijke registratiesystemen; de bruikbaarheid daarvan binnen het voorgestelde onderzoeksdesign kent enkele beperkingen, die in kaart gebracht worden. In het onderzoek wordt als eerste een selectie gemaakt van goed en minder goed presterende scholen, al dan niet met onderwijsachterstandenbekostiging. Vervolgens worden deze scholen kwalitatief onderzocht op succes- en faalfactoren, waarna in een quasi-experimentele opzet op

slechter presterende scholen deze effectieve interventies en succesfactoren worden geïmplementeerd of versterkt. Zie hieronder een opsomming van deze interventies en succesfactoren.

Schema – Overzicht effectieve interventies en succesfactoren
(V)=vve, (S)=onder schooltijd, (A)=*afterschool*

Effectieve interventies:	
• vve (o.a. Piramide, Kaleidoscoop)	V
• schakelklassen	A, S
• verlengde leertijd, zomerschool, buitenschoolse programma's	A
Succesfactoren:	
• kwaliteit van (de uitvoering van) het programma	V, S, A
• pedagogiek: focus op leren en onderwijzen	V, S, A
o instructie gericht op leerprestaties; professionalisering (*learning community*), structuur, beloning, uitdagingen bieden, hoge verwachtingen hebben van leerlingen	V, S, A
o toetsing	S
o kwalitatief hoogwaardig, praktijkgericht, geïntegreerd curriculum (waaronder kunstonderwijs)	S, A
• ouderbetrokkenheid ('schooltaal' leren verstaan en kinderen ondersteunen bij schoolwerk; psychosociale omgeving en intellectuele stimulering thuis; informatievoorziening en ondersteuning; actieve benadering van kind)	V, S, A
• schoolleiderschap (procesgericht, *collaborative* leiderschap*)*	V, S
• hooggekwalificeerde medewerkers (algemeen, programmaspecifiek)	V, A
• participatiegraad (al dan niet vaak de activiteiten bezoeken)	V, A
• middellange duur (44-210 uren)	V, A
• constructieve schoolcultuur (open communicatie, samenwerking, gering verloop onder leraren, perceptie van leerkrachten op de effectiviteit van een interventie, familiegevoel op school, empathisch contact leraar-leerling)	S, A
• school- en klasomstandigheden (schaalverkleining, *compositional effect,* fysieke voorwaarden; klassenmanagement, veilige ruimte)	S, A
• thuistaal is Nederlands	V
• doorgaande lijn van voor- naar vroegschoolse periode (van peutergroep naar groep 1) en naar groep 3 en later	V
• kleine groepen en dubbele bezetting (gunstige kind-staf-ratio)	V
• goede mix van kinderen met diverse sociaal-culturele achtergronden	V
• centrumgericht programma (continuïteit), liefst in combinatie met gezinsgericht programma	V
• modelprogramma's met bij voorkeur huisbezoek	V
• informatierijke omgeving (datagestuurd onderwijs, toetsen voor leerlingen, tevredenheidsonderzoek onder leerlingen en personeel)	S
• adequate mate van financiering	S
• externe support (netwerken en monitoring)	S
• schoolgerelateerde extracurriculaire activiteiten, hoewel anderen juist niet-schoolgerelateerde (verrijkings)activiteiten aanbevelen	A
• prestatiegraad (laagpresteerders hebben er meer aan dan cognitief zeer succesvolle leerlingen)	A
• afgestemd op leeftijdsgroep	A
• mix van interventies (geen eenzijdigheid)	A
• één-op-één tutoring	A

1 Achtergronden van het onderzoek

1.1 Inleiding

Onderwijsachterstandenbeleid staat in Nederland erg in de belangstelling. Niet alleen is er veel overheidsgeld mee gemoeid, maar ook de persoonlijke levensloop en (on)mogelijkheden van individuen zijn in het geding. En daarmee is het dan ook meteen een maatschappelijk probleem. Ledoux *et al*. (2011) formuleren het hoofddoel van onderwijsachterstandenbeleid:

> "Het hoofddoel van het onderwijsachterstandenbeleid is vermindering van de ongelijkheid in schoolprestaties die samenhangt met groepskenmerken, met name sociaal milieu en etnische herkomst." (p. 5)

Rond het jaar 2000 is dit algemene doel nader omschreven als:

- verbeteren van de startpositie van doelgroepleerlingen bij binnenkomst in de basisschool;
- terugdringen van de taalachterstand bij de doelgroepleerlingen;
- verhogen van de doorstroom van deze leerlingen naar hogere vormen van voortgezet onderwijs (Ledoux *et al*., 2011).

Niet kunnen meekomen op school als gevolg van het behoren tot een bepaalde maatschappelijke groep die door omstandigheden minder kans heeft op schoolsucces, heeft vergaande persoonlijke en maatschappelijke consequenties. Vanwege het streven naar gelijke kansen in het onderwijs is het van belang onderwijsachterstanden in een vroeg stadium te voorkomen. In Nederland is daartoe voor- en vroegschoolse educatie (vve) in het leven geroepen (Ledoux *et al*., 2011). Desondanks kunnen er op de basisschool toch nog onderwijsachterstanden aanwezig zijn of zelfs ontstaan. In dat geval kan een groep leerlingen tijdens schooltijd hiervoor bijgespijkerd worden. Ook na de reguliere schooltijd zijn er mogelijkheden om opgelopen leerachterstand te bestrijden. Uitgangspunt is dat talenten van leerlingen het meest waardevolle kapitaal zijn van onze samenleving. Prestaties in het primaire onderwijs, op basis van de aanwezige talenten, dienen te worden gestimuleerd, zodat zij een sterke basis vormen voor het vervolgonderwijs en uiteindelijk bijdragen aan een gelukkige en economisch welvarende samenleving. De ambities binnen het primair onderwijs zijn beschreven in verschillende documenten: Actieplan Basis voor Presteren (PO), Actieplan Leraar 2020-een krachtig beroep!, Kennisagenda OCW en zeer recent het Bestuursakkoord Primair Onderwijs 2012-2015.

In de voorbije jaren zijn tal van beleidsmaatregelen ingezet om de verschillen tussen groepen leerlingen te nivelleren. De oudste en naar financiële maatstaven de belangrijkste is de gewichtenregeling. Dit is een regeling die voorziet in extra geld voor scholen met een substantieel aantal doelgroepleerlingen. Doelgroepleerlingen tellen in de formatieberekening extra zwaar mee en dit levert de school dus extra middelen op. De gewichtenregeling heeft in de loop der jaren enkele wijzigingen ondergaan (Ledoux *et al*., 2011).

Uit descriptief onderzoek van Ledoux *et al*. (2011) blijkt dat allochtone doelgroepleerlingen in het primair onderwijs een deel van hun achterstand inlopen, maar het is onbekend of deze winst aan het onderwijsachterstandenbeleid kan worden toegeschreven. Daarvoor is het inzicht in (de effectiviteit van) interventies (nog) te beperkt.

Over het algemeen kan gesteld worden dat er een beperkt inzicht is in interventies op het gebied van onderwijsachterstandenbestrijding en de effectiviteit ervan.

1.2 Onderzoeksvragen

In dit rapport wordt nader ingegaan op de volgende centrale onderzoeksvraag:

Welke door scholen gepleegde interventies dragen bij aan een effectieve bestrijding van onderwijsachterstanden in het primair onderwijs?

Uit de centrale onderzoeksvraag vloeien de volgende onderzoeksvragen voort:

1. welke interventies komen als effectief naar voren in internationale wetenschappelijke literatuur?
2. op welke wijze werd de effectiviteit van interventies onderzocht en wat is de bewijskracht van deze onderzoeken?
3. welke valkuilen, voor- en nadelen zijn er bij het onderzoeken van de effectiviteit?
4. hoe kan de effectiviteit van interventies worden onderzocht?

1.3 Methode

De in § 1.2 genoemde onderzoeksvragen worden met behulp van literatuurstudie beantwoord. Via sEURch, een zoekmachine van de Erasmus Universiteit Rotterdam, zijn onder andere de databases van ERIC, PiCarta, PsychINFO, CCT, Science Direct, Sociological Collection, Google Scholar en WorldCat benaderd. In de zoektocht naar relevante onderzoeksliteratuur is de volgende zoekstrategie (trefwoorden en hun specifieke combinaties) aangehouden:

Aspect 1

Inequality of education (educational inequality), urban education, inner-city schools, policy, urban schools, disadvantaged (students), disadvantaged schools, dropouts, ethnic minorities, low socio economic background, school failure, segregation;

Aspect 2

Student performance, student achievement, student motivation, learning performance, learning skills, competency, accomplishments;

Aspect 3

Primary schools, primary education;

Aspect 4

Evaluation standards, educational program performance, program effectiveness, intervention, school effectiveness, school improvement;

Aspect 5

Review, meta-analysis, meta-study;

Aspect 6

Publicatiedatum vanaf 2000.

De per aspect genoemde woorden zijn in meer of mindere mate synoniemen van elkaar. Met deze trefwoorden is via de OR-operant gezocht. De zes aspecten zijn met elkaar gecombineerd via de AND-operant, zodat alle bronnen in principe voldoen aan alle zes aspecten. Met de toevoeging van het aspect ‘review, meta-analysis, meta-study’ werd beoogd een selectie van artikelen te krijgen die de onderzoeksliteratuur samenvatten (zoals meta-analysen, systematische en verhalende (*narrative*) reviews). Hierdoor hoefden de onderzoekers niet zelf een literatuurreview te maken, maar de bestaande reviews op de verschillende gebieden van onderwijsachterstanden samen te vatten. De belangrijkste reden hiervoor is kosten- en tijdefficiëntie. Eigenstandig onderzoek werd alleen ‘toegelaten’ wanneer het van zeer recente datum was (2011 of 2012), omdat de kans dat dergelijk materiaal is verwerkt in gepubliceerde reviews klein is. Voor secundaire analysen geldt hetzelfde.

Buitenlandse literatuur kent veelal niet het blijkbaar nogal Nederlandse onderwijsachterstandenbeleid. In de synoniemen (aspect 1) is een parallel getrokken met verschijnselen die met onderwijs-achterstanden samenhangen. Uit de literatuur blijkt onderwijsachterstand vaak samen te hangen met lage sociaaleconomische status en etniciteit. Door deze handelwijze impliceren de onderzoekers niet dat alle kinderen die hieraan voldoen, minder op school zouden kunnen meekomen en, omgekeerd, dat leerlingen die hieraan niet voldoen, niet te maken kunnen hebben met onderwijsachterstanden.

Er werd bij de keuze van zoektermen op gelet geen *voorbeelden* van mogelijk effectieve interventies bij onderwijsachterstanden te gebruiken, omdat dat de *output* moet zijn van de searches en niet de *input* aan zoektermen.
Referenties uit artikelen, met name reviews, hebben nog weer nieuwe artikelen opgeleverd (sneeuwbalmethode).
Daarnaast is een zoektocht ondernomen naar relevante internationale websites waarin informatie is opgenomen over effectiviteit van het onderwijsachterstandenbeleid. Voorbeelden hiervan zijn de Amerikaanse website *What Works Clearinghouse* (WWC) en de Britse website van CARE (*Centre for Applied Research in Education*). Daarnaast zijn er websites gescreend met informatie over evaluaties van onderwijsbeleid en overzichtsverslagen van beleidsgeoriënteerde projecten (OECD / UNESCO).

Globale beschrijving van de onderzoeksliteratuur
Door globale bestudering van de samenvattingen werd beoordeeld of een artikel of rapport mogelijk interessant kon zijn om één van de onderzoeksvragen te beantwoorden. Er is in databanken gezocht naar reviews en meta-analysen op het gebied van het bestrijden van onderwijsachterstanden. Recente onderzoeksartikelen of rapporten gepubliceerd in 2011 of 2012 zijn ook bij de selectie betrokken. Op grond hiervan werden 86 artikelen en rapporten geselecteerd voor nadere bestudering. Bij grondiger lezing viel op dat toch een flink aantal bronnen niet over onderwijsachterstanden ging, maar schoolprestatieverbetering in het algemeen. De focus van de onderzoeken in deze reviews was vaak niet gericht op het primair onderwijs maar betrof *early childhood* of vroeg- en voorschoolse educatie (vve). Toch is besloten die laatste serie bronnen wel te bestuderen, waarbij vve als interventie wordt gezien ter voorkoming van à priori onderwijsachterstanden in het basisonderwijs. In een flink aantal bronnen werd geen specifiek onderscheid gemaakt naar primair onderwijs, ook al omdat bijvoorbeeld in de Verenigde Staten de *grade*-systematiek niet correspondeert met het Nederlandse onderwijssysteem. Een aantal artikelen viel af omdat de uitkomstmaat louter niet-cognitief was, waartoe de opdracht zich echter wel beperkte.
Al met al bleven er 33 literatuurbronnen over. Hiervan hebben er twaalf betrekking op preschools, tien gaan over verschillende vormen van *afterschooling* en zes over (onder andere) het basisonderwijs. De resterende vijf bronnen zaten in de aanvankelijke selectie, werden inhoudelijk geëxcludeerd, maar bevatten wel relevante opmerkingen over onderzoeksdesigns.
Verreweg de grootste groep artikelen (22 studies) is te bestempelen als *narrative* literatuurreviews, al dan niet als uitgebreide inleiding op een eigenstandig onderzoek. De search leverde vijf systematische meta-analysen op. Onderzoek dat vanwege de recente publicatiedatum meegenomen werd, betrof twee quasi- experimenten, één correlationeel onderzoek en één secundaire analyse. Tenslotte zaten bij de geselecteerde literatuur twee methodenevaluaties en een uitgebreide Nederlandse beleidsbrief.
De herkomst van de literatuur is in dit geval minder interessant omdat het met name gaat om verhalende of systematische reviews, waarbij internationaal naar bronnen is gezocht. Een Duitse review auteur beschrijft dus niet de Duitse situatie, maar woont toevallig in dat land.
Als speciale bron dient toch wel het standaardwerk van Hattie (2009) genoemd te worden. Hattie bundelt in de 2009-versie meer dan 800 reviews op het gebied van onderwijseffectiviteit, waarvan hij in 2012 nog weer een update heeft verzorgd (Hattie, 2012).
Via de sneeuwbalmethode zijn de overige bronnen achterhaald.

1.4 Opbouw van het rapport

De literatuur die behulpzaam is bij de beantwoording van de onderzoeksvragen laat zich grofweg in drie categorieën verdelen. De eerste categorie betreft interventies die beogen te voorkomen dat er onderwijsachterstanden zullen ontstaan in het primair onderwijs of deze à priori proberen te verminderen. Dergelijke interventies liggen op het gebied van *preschooling*, hetgeen in de Nederlandse situatie voorschoolse educatie heet, als onderdeel van voor- en vroegschoolse educatie (vve).
Een tweede groep interventies betreft maatregelen die genomen worden gedurende de reguliere basisschooltijd. In tegenstelling tot vve, dat feitelijk één interventie is met verschillende verschijningsvormen, is deze groep interventies meer divers van aard.

Het derde cluster interventies, dat zich in de onderzoeksliteratuur aftekent, is eigenlijk ook weer één interventie, met uiteenlopende verschijningsvormen. Het gaat hierbij om verschillende wijzen van *afterschooling*.
De onderzoeksliteratuur wordt beschreven in *hoofdstuk 2 tot en met 4*, aan de hand van bovenstaande drieslag. In wezen is deze drieslag met het oog op het primair onderwijs samen te vatten als interventies uit te voeren vóór, tijdens of na de reguliere basisschooluren.
Hoofdstuk 5 gaat verder in op welke wijze effectiviteit van interventies is aangetoond. De voor- en nadelen van verschillende, in de praktijk gebruikte designs worden samengevat zoals die door andere onderzoekers in hun artikelen zijn genoemd. Dit mondt uit in een schets van mogelijke onderzoeksdesigns. De opdrachtgever heeft in een voorbespreking een keuze gemaakt, die vervolgens verder is uitgewerkt door de onderzoekers.
In *hoofdstuk 6* worden de bevindingen bediscussieerd.

2 Onderzoek interventies vóór reguliere basisschooltijd: preschool en vve

2.1 Inleiding

Voorschoolse en vroegschoolse educatie (vve; in het Engels vaak '*preschool*' genoemd, hoewel dat iets beperkter van opzet lijkt en slechts de peuterleeftijd omvat) is een middel om onderwijsachterstanden te bestrijden. Geddes *et al*. (2011) stellen dat *preschooling*-programma's bedoeld zijn ter voorkoming van onderwijsachterstanden en zich vooral richten op *at risk* kinderen. Meijnen (2004) stelt dat onderwijsachterstandenbeleid tien jaar geleden vooral gericht was op de basisschoolleeftijd, maar dat dit niet tot bevredigende resultaten leidde, omdat in groep 2 al grote verschillen tussen kinderen bestonden, waarbij vooral allochtone kinderen in het nadeel zijn en met een grote achterstand begonnen aan het taal- en leesonderwijs. Voor- en vroegschoolse educatie is onderwijs voor zeer jonge kinderen met een (taal) achterstand. Hiermee kunnen kinderen spelenderwijs hun taalachterstand inhalen, zodat zij een goede start kunnen maken op de basisschool. Nederland is hierin niet uniek. Ook andere landen bieden voorschoolse educatie om achterstanden te bestrijden (OECD, 2006) Voorschoolse educatie is bedoeld voor peuters van 2,5 en 3 jaar en vindt plaats op de peuterspeelzaal of de kinderopvang. Gemeenten zijn verantwoordelijk voor de voorschoolse educatie en bepalen welke kinderen hiervoor in aanmerking komen. Vroegschoolse educatie wordt gegeven in groep 1 en 2 van de basisschool. Voor vroegschoolse educatie zijn basisscholen verantwoordelijk. Vve-doelgroepkinderen zijn een grote groep kinderen. Zo behoort in Amsterdam ongeveer de helft van alle twee- tot en met vijfjarigen tot deze groep (Inspectie van het Onderwijs, 2008). Er zijn diverse programma's voor vve. Ze richten zich vooral op taalachterstanden en rekenen, maar besteden ook aandacht aan de sociaal-emotionele ontwikkeling en de motoriek van kinderen. In de Regeling Voor- en vroegschoolse educatie (OCW, 2000), het Besluit Vaststelling doelstelling en bekostiging onderwijsachterstanden 2006-2010 (OCW, 2006) en Besluit specifieke uitkeringen gemeentelijk onderwijsachterstandenbeleid 2011-2014 (OCW, 2010), wordt niet aangegeven welke vve-programma's moeten worden aangeboden en welke programma's de voorkeur verdienen. De gemeenten kiezen zelf programma's van waar uit gewerkt wordt. Programma's kunnen ontwikkelingsgericht zijn en kindvolgend met ruimte voor eigen initiatief van het kind of meer programmagestuurd met een voorgestructureerd inhoudelijk programma.
Op 1 augustus 2010 is de wet OKE (Ontwikkelingskansen door Kwaliteit en Educatie) in werking getreden. Hierin wordt aangestuurd op meer en betere voorschoolse educatie in de kinderopvang en peuterspeelzalen. Kinderen met een taalachterstand kunnen zo beter worden geholpen om die achterstand weg te werken. Peuterspeelzalen en kinderopvanginstellingen gaan hiervoor meer samenwerken. Het minimumniveau van de kwaliteitseisen aan peuterspeelzalen komt meer in lijn met dat van de kinderopvang.
Gemeenten hebben de wettelijke verantwoordelijkheid om een goed voorschools aanbod te doen aan alle jonge kinderen met een taalachterstand. De gemeenten hebben een inspanningsverplichting om bijvoorbeeld in samenwerking met de consultatiebureaus al deze kinderen te bereiken.
Onderdeel van de wet OKE is een landelijk kwaliteitskader voor peuterspeelzalen. Op een groep van maximaal zestien kinderen staan twee pedagogisch medewerkers, waarvan minimaal één met een opleiding op SPW-3-niveau of daaraan gelijkwaardig. Als er voorschoolse educatie wordt gegeven, moeten er twee beroepskrachten zijn met een goede beheersing van de Nederlandse taal. In de komende jaren wordt ernaar gestreefd ook het aantal hbo-gekwalificeerde begeleiders op vve-locaties te verhogen. Alle gemeenten moeten voldoen aan de eisen uit de Wet OKE. In kleinere gemeenten (G37) is het vaak moeilijker om een aanbod te realiseren en doelgroepkinderen te bereiken dan in grotere gemeenten (Bijl *et al.*, 2011). De GGD houdt toezicht op de kwaliteitseisen in peuterspeelzalen. De Inspectie van het Onderwijs houdt toezicht op de voor- en vroegschoolse educatie.

De vraag waarover dit hoofdstuk gaat, is wat de laatste tien jaar heeft gewerkt om achterstanden te bestrijden en hoe de werkzame ingrediënten van vve-programma's het beste onderzocht kunnen worden in een goed design.

Er zijn het afgelopen decennium diverse studies en meta-analysen gepubliceerd waarin een gevarieerd beeld naar voren komt over de effectiviteit van vve-programma's. Soms zijn er medium effectgrootten gevonden met betrekking tot de leerprestaties van kinderen; soms zijn de effecten klein of geheel afwezig.

2.2 Resultaten

Achtereenvolgens worden in deze paragraaf studies beschreven die licht werpen op effecten van vve of *preschooling* respectievelijk op het ontbreken daarvan.

Positieve effecten van vve
Het onderzoek van Veen *et al*. (2000) heeft de aanzet gegeven voor de vve-regelingen. In dit onderzoek werden kinderen die aan het programma Piramide of Kaleidoscoop deelnamen, vergeleken met kinderen uit een controlegroep zonder programma. Kinderen werden niet random toegewezen maar gematcht (quasi-experimentele opzet). Ook is er onderscheid gemaakt tussen kinderen die vanaf het begin het aanbod kregen en kinderen die later instroomden. In de experimentele groepen werd voldaan aan de ideale condities van een dubbele bezetting en een aanbod van vier dagdelen per week. De analysen zijn gebaseerd op resultaten van 117 kinderen. Er werden positieve effecten gevonden voor de cognitieve en taalontwikkeling van de kinderen met een gemiddelde effectgrootte, maar geen duidelijke effecten voor de sociaal-emotionele ontwikkeling.
Kinderopvang en peuterspeelzalen maken kinderen taliger, verbaler en sociaal vaardiger. Dit wordt gesuggereerd door auteurs uit meerdere Nederlandse studies zoals Groot *et al*. (2001), Leseman (2002) en de review van Van der Vegt *et al*. (2007). Over het algemeen wordt dan ook gesteld dat vroege educatie een positief effect heeft op de ontwikkeling van jonge kinderen.
Onderzoek van Steensel *et al*. (2007) wijst in de richting dat het bezoeken van de peuterspeelzaal een gunstig effect heeft op de talige ontwikkeling van kinderen. De onderzochte peuterspeelzalen hadden nog geen vve-programma ingevoerd. Kinderen die op een voorschoolse voorziening hebben gezeten, blijken in groep 2, 3 en 4 van de basisschool significant hogere scores te halen op de Cito-toetsen voor woordenschat, technisch en begrijpend lezen. Andere effecten worden gevonden aangaande de kind/leidsterratio: hoe minder kinderen per leidster des te beter. Als er thuis doorgaans Nederlands wordt gesproken, is het effect van deelname aan een voorschoolse voorziening groter.
Er zijn meer aanwijzingen voor positieve effecten van voorschoolse voorzieningen. Leseman (2007) stelt in een beleidsartikel (*Early Education for Immigrant Children*) op basis van een aantal reviews (waar onder Blok *et al*., 2005) dat een *center-based* benadering zoals in de voorschool gebeurt, het meest effectief is in combinatie met interventies om ouders te betrekken, te informeren en te ondersteunen. Wanneer er effecten worden gevonden, benadrukken auteurs dat de kwaliteit van de voorzieningen van doorslaggevend belang is. Deze kwaliteit is niet altijd vanzelfsprekend. Zo blijkt ook uit een onderzoek van het Nederlands Consortium Kinderopvang Onderzoek (Vermeer *et al*., 2005; Gevers Deynoot-Schaub *et al*., 2009) een landelijke peiling onder kinderdagverblijven, dat de proceskwaliteit (waarbij het gaat om gezondheid, hygiëne, veiligheid, aanwezigheid van gevarieerd ontwikkelingsmateriaal) was gedaald ten opzichte van voorgaande peilingen. Ook de pedagogische kwaliteit kan beter. Deskundigheidsbevordering is hierbij belangrijk zodat de interactievaardigheden van de pedagogisch medewerkers in de toekomst verbeteren (De Kruijf *et al*., 2009).
Van Schooten (2012) schrijft een verhandeling over cognitieve leerprestatieverbetering als gevolg van voorschoolbezoek. Hij betrekt hierin diverse bronnen, waaronder meta-analysen. Hij concludeert dat vve met name voor onderwijsachterstandsleerlingen effectief is. De interactie tussen leidster en kind, goede training voor de leidster in het gebruik van het programma alsook lange interventieduur en kleinschaligheid zijn enkele van de elementen die als voorwaarden voor effectiviteit aangewezen kunnen worden.

Ook studies in het buitenland laten positieve effecten van *preschooling* zien. In de meta-analyse van Anderson *et al*. (2003) worden negen studies besproken waarbij effecten worden besproken van *early childhood development* op cognitieve prestaties, plaatsing in speciaal onderwijs, sociaal gedrag, gezondheid en *family outcomes* waarbij het volgen van hoger onderwijs, het hebben van werk en

gezinsinkomen zijn meegenomen. In zes studies worden positieve effecten gevonden, één studie rapporteert een negatief effect en in twee studies is het effect onduidelijk.
Een studie die veel wordt aangehaald, is die van de Nobelprijswinnaar voor de economie James Heckman die op basis van langdurig onderzoek positieve effecten van *early education and care* vaststelt: betere schoolprestaties, minder misdaad, meer kans op werk en minder uitkeringsafhankelijkheid. Heckman (2000) concludeert dat vve één van de beste investeringen is die de overheid kan doen. Vve levert het grootste effect op bij kinderen in achterstandsituaties. Ook in de studie van Wolfe & Tefft (2007) worden goede effecten van interventies vanuit een economische insteek gerapporteerd. De scope van risicokinderen is in deze review breed gedefinieerd, ook studies naar kinderen met een laag geboortegewicht en kinderen met crimineel gedrag zijn hierbij betrokken. Volgens de auteurs waren van de achttien studies de uitkomsten van zes studies 'relevant' waarmee bedoeld is dat er een aantal significante positieve uitkomstresultaten zijn op een aantal brede gebieden en dat deze positieve uitkomsten ook in een follow-up terug te vinden zijn; negen studies leverden minder relevante resultaten op en drie waren niet relevant. Ook zijn goede resultaten van vroege interventie voor de cognitieve en sociale vaardigheden van jonge kinderen gevonden door Peisner-Feinberg (2001). In een meta-analyse van Burger (2010) worden de evaluatieonderzoeken van 23 vroeg- en voorschoolse projecten uit de hele wereld besproken. In de meeste studies worden kleine tot middelgrote effecten gevonden op ontwikkelingsdomeinen als woordenschat, taal, lezen en rekenen/wiskunde. Burger vindt het riskant om conclusies te trekken vanwege de diversiteit aan studies en de wisselende uitkomsten, maar is van mening dat de meeste programma's positieve effecten opleveren.
Geddes *et al*. (2011) verrichtten een literatuurstudie naar de cognitieve taalontwikkeling van 0–5-jarigen. Als kwalitatief hoogwaardige *early childhood education* gecombineerd wordt met een paar huisbezoeken om het leren in de thuissituatie te kunnen verbeteren, dan heeft dat gunstige effecten op de leerprestaties van een kind. Net als Blok *et al*. (2005) vinden zij dat een combinatie van *center*- en *homebased services* het meest effectief zijn, die dan bovendien gericht zijn op ouder èn kind.
Kinderen van laag opgeleide ouders hebben grote kans een achterstand op te lopen, nog vóór ze naar het basisonderwijs gaan. Biedinger (2011) ontdekte in haar secundaire analyse van cognitieve vaardigheden van 3–4-jarigen dat actieve bemoeienis van ouders met hun kinderen een eventueel laag ouderlijk opleidingsniveau kan compenseren. Als ouders die actieve aandacht niet kunnen geven, is volgens Reynolds *et al*. (2006) de *kindergarten* een effectief alternatief dat kan compenseren voor een weinig uitdagende thuissituatie. Achtergestelde (*disadvantaged*) kinderen hebben het meeste baat bij kwalitatief hoogwaardige programma's. Ze beschrijven vier programma's die effectief blijken, maar wat nu precies de effectieve elementen zijn, blijft onbekend. Als van belang zijnde factoren worden aangeduid:

- een focus op talige voorbereiding voor de basisschool (*schoolreadiness*);
- een langdurig programma (meerdere jaren);
- vanuit de school georganiseerd vanwege de continuïteit;
- ouderbetrokkenheid;
- inzet van klassenassistenten en klassenverkleining, waardoor de ratio leerkracht/leerling kleiner wordt en er meer aandacht is voor de leerling.

De kwaliteit van de onderzoeksdesigns is wisselend. Als er al een gerandomiseerde, gecontroleerde studie wordt uitgevoerd, waarbij kinderen random aan een interventie worden toegedeeld, is dat doorgaans een studie met een beperkt aantal kinderen. Een voorbeeld hiervan is de studie van Schweinhart *et al*. (2005). In deze studie zijn langdurige, positieve effecten gevonden van deelname aan de *High/Scope Perry Preschool* die tot in de volwassenheid meetbaar zijn (volgen van hoger onderwijs, minder criminele carrières, minder tienerzwangerschappen en hogere salarisschalen). Het project is geëvalueerd in de gerandomiseerde gecontroleerde studie waarbij 58 kinderen random werden toegewezen aan het programma en 65 kinderen niet. In de *High/Scope Perry Preschool* zijn hoog gekwalificeerde leerkrachten werkzaam en is er een laag kind/leidsterratio. Dit programma is overigens niet nationaal uitgerold.
Burger refereert aan Reynolds *et al*. (1997) die stellen dat het empirische bewijs vooral afkomstig is van modelprogramma's, waarbij programma's onder optimale condities plaatsvinden. Burger concludeert voorzichtig dat de meerderheid van de studies laat zien dat vroege interventie kinderen een

betere start geeft op school. Kortetermijneffecten zijn eenvoudiger te vinden dan langetermijneffecten. Langetermijneffecten zijn schaars, maar soms worden deze effecten wel gevonden zoals in de studie van Schweinhart *et al*. (2005). Harrison *et al*. (2012) bespreken in hun review dat kinderen, met name op het gebied van taal en geletterdheid, zeker kunnen profiteren van een goede kwaliteit van voor- en vroegschoolse educatie, maar dat nog onduidelijk is welke onderdelen de werkzame ingrediënten zijn.
Langetermijneffecten van voor- en vroegschool worden gerapporteerd in studies van Reynolds *et al*. (2007, 2011) onder *low-income families*. In deze studies worden meer dan 1.000 participanten gevolgd tot in hun volwassenheid. Deelname van kinderen uit *low-income families* aan een voor- of vroegschool correleert positief met een hogere opleiding en hoger inkomen op latere leeftijd.
Reynolds *et al*. (2006) en Van Schooten (2012) waarschuwen echter ervoor dat effecten van *preschooling* de neiging hebben uit te doven.
In Engeland zijn er positieve resultaten te zien in het project *Effective Provision of Preschool Education* (EPPE, 2004). Het EPPE-project is de eerste Europese longitudinale studie naar de ontwikkeling van kinderen (3 tot 7 jaar) uit een grote, representatieve steekproef van 3.000 kinderen. De effecten van *preschool education* van 3- en 4-jarigen zijn onderzocht. Ook zijn er gegevens verzameld over de omgeving thuis. De kinderen waren afkomstig uit 141 organisaties voor peutereducatie en speelgroepen. De kinderen uit de experimentgroep werden vergeleken met 'thuiskinderen'. In het EPPE-project zijn er twaalf intensieve case-studies verricht naar de karakteristieken en onderliggende pedagogische beginselen van succesvolle settings met goede leerlinguitkomsten.
De bevindingen zijn als volgt:

- het bezoeken van een *preschool* is goed voor de cognitieve en sociale ontwikkeling van het kind;
- een fulltime programma is niet per se beter dan parttime programma;
- langetermijneffecten tot zeven jaar zijn zichtbaar en zijn sterker naarmate kinderen langer naar een voorschoolse voorziening zijn gegaan;
- *preschool* is met name goed voor doelgroepkinderen als er een goede mix is van kinderen met diverse sociaalculturele achtergronden;
- kinderopvang geïntegreerd met de school levert de beste resultaten op;
- de kwaliteit van de *preschool* is gerelateerd aan de ontwikkelingsuitkomsten van kinderen;
- kinderen maken meer vooruitgang als de professionals in de *preschool* hoger zijn opgeleid of beter getraind;
- als zowel leren als sociale ontwikkeling aandacht krijgen, levert dat de beste uitkomsten op;
- het bieden van een rijke leeromgeving waarin kinderen worden gestimuleerd om te ontdekken en na te denken, levert de beste resultaten op;
- een goede, stimulerende leeromgeving thuis is belangrijk. Dit hangt voor een deel samen met sociale klasse. Het opvoedend gedrag van ouders is belangrijker dan hun sociaal milieu.

Isaacs (2008) wijst programma's aan die in het buitenland model staan en goede resultaten opleveren met betrekking tot de schoolprestaties van kinderen. Door deze modelprogramma's bezoeken leerlingen in mindere mate het speciaal onderwijs, blijven niet zitten en hebben hogere IQ-scores. De modelprogramma's zijn intensief en langdurig:

- the *High/Scope Perry Preschool Project* in Ypsilanti, Michigan (Schweinhart *et al*., 2005); dit is een intensief tweejarig programma voor 3- en 4-jarigen, waarbij er wekelijks een huisbezoek plaatsvindt; de ratio kind/leidster is 6:1; de kosten bedragen per kind ruim 14.000 dollar voor deelname aan het programma;
- the *Abecedarian Project* in Chapel Hill, North Carolina; dit is een zeer intensief en duur programma van vijf dagen per week voor peuters en kleuters gedurende meerdere jaren; de ratio kind/leidster voor peuters is 3:1 en voor kleuters 6:1; ook huisbezoeken horen drie jaar lang bij het programma; de kosten per kind zijn meer dan 40.000 dollar voor deelname aan het programma;
- the *Chicago Child-Parent Centers Program* in Chicago, Illinois; kinderen nemen halve dagen deel aan het programma; daarnaast is er ook een intensief ouderprogramma en een zes weken durend zomerprogramma; de kosten per kind bedragen bijna 7.000 dollar voor deelname aan het programma.

Andere goede voorbeelden zijn:

- the *Turkish Early Enrichment Program (TEEP)* in Istanbul; een centrumgericht programma in combinatie met een gezinsgericht programma voor *low-income families*;
- the *Syracuse Family Development Research Project* in Syracuse, New York;
- the *Yale Child Welfare Project* in New haven, Connecticut.

Niet zonder reden stelt Leseman (2007) dan ook dat alleen aanbod van bovengemiddelde kwaliteit in kinderdagverblijven kan zorgen voor een meetbaar betere ontwikkeling (medium effectgrootte) van doelgroepkinderen op cognitief en talig gebied. Een intensieve aanpak in *early education* lijkt het meest succesvol in de review van Gorey (2001). Bij minder intensieve programma zijn de effecten geringer. De wijze waarop vve wordt aangeboden is essentieel. De resultaten kunnen substantieel zijn en langdurig (met een medium effectgrootte) als het vve-aanbod kwalitatief hoogstaand is en de kinderen een lange periode een intensief en goed programma krijgen.
In voor- en vroegschoolse programma's zijn er goede ervaringen met combinatieprogramma's waarbij ook ouders betrokken worden. De effecten zijn niet als vanzelf groot als er een ouderprogramma is, maar wel als ouders gestimuleerd worden hoge aspiraties en verwachtingen te hebben en een actieve benadering te kiezen om hun kind te laten leren. Er zijn hoge effectscores als ouders hun kind laten voorlezen (d=.51), hoger dan wanneer ouders zelf voorlezen (d=.18). Deze effecten gelden voor kinderen met en zonder leesproblemen en voor gezinnen met lage en gezinnen met hoge *SES*.
In een evaluatiedesign mogen, tenslotte, de volgende onderdelen niet ontbreken (Leseman, 2007): (a) de kwaliteit van de vve, (b) de inhoud van de interventies en (c) de continuïteit van de interventie, vertaald in de integratie met een programma voor ouders, samenwerking met andere instellingen en een goede aansluiting met het primair onderwijs.

Naast de positieve geluiden over de effecten van voor- en vroegschoolse educatie bestaan er ook veel onderzoeken waarin deze effecten niet kunnen worden aangetoond.

Geen effecten van vve
Het Sociaal en Cultureel Planbureau heeft in 2003 op basis van het PRIMA-cohortonderzoek studie gemaakt van de effecten van voorschoolse voorzieningen.
Uit het SCP-rapport (Dagevos *et al*., 2003) blijkt dat kinderdagverblijven geen positief effect hebben op de ontwikkeling van kinderen. Er is zelfs een licht negatief effect. Destijds hadden kinderdagverblijven zelden een sterke pedagogische aanpak. Inmiddels is dat sterk verbeterd en hanteert elk kinderdagverblijf een pedagogisch beleidsplan. Uit hetzelfde onderzoek blijkt overigens wel dat er een klein positief effect is van het bezoeken van peuterspeelzalen op de cognitieve ontwikkeling van het kind.
Volgens de Onderwijsraad (2002) is er over de effecten van peuterspeelzalen weinig bekend. Er wordt gesignaleerd dat peuterspeelzalen onvoldoende zijn toegerust om te voldoen aan de gewenste kwaliteitsaspecten. De kwaliteit van het aanbod en de deskundigheid van de leidsters lijken de belangrijkste knelpunten volgens de MO-Groep (2001).
Het effect van vve-programma's in Nederlands onderzoek is niet overweldigend (Veen *et al*., 2006). Uit de opsomming van onderzoek uit het rapport van Van der Vegt (2009) blijkt dat niet aangetoond kan worden dat vve in de vier grote steden werkt.
Uit het onderzoek van Veen *et al*. (2008) in Rotterdam blijkt dat deelname aan een reguliere speelzaal of crèche dezelfde resultaten oplevert als deelname aan een vve-programma van de peuterspeelzaal. Ook een ander onderzoek van het SCO-Kohnstamm Instituut uit 2009 (Van Schooten & Sleegers, 2009) naar de effectiviteit van vve in Oosterhout en Den Bosch levert geen positieve effecten op met betrekking tot de cognitieve en sociaal emotionele ontwikkeling van kinderen.
Het programma Startblokken is in 2006 onderzocht door het SCO-Kohnstamm Instituut op dezelfde wijze zoals dat destijds is gedaan door Veen *et al*. (2000) voor de programma's Piramide en Kaleidoscoop. Voor Startblokken konden geen positieve effecten gevonden worden op de cognitieve ontwikkeling van kinderen. In de meta-analyse van Burger (2010) worden bij enkele studies geen effecten gevonden voor voor- en vroegschoolse voorzieningen op onder andere de taal- en rekenprestaties op latere leeftijd zoals in de studie van Driessen (2004). Kinderen die een peuterspeelzaal bezochten scoorden hoger op taal- en rekentoetsen dan kinderen die dat niet deden. Bij

correctie voor achtergrondkenmerken zoals opleiding ouders, etniciteit en geslacht, verdwenen deze samenhangen.
Uit de effectanalysen van Nap-Kolhoff *et al*. (2008) van een onderzoek onder 253 scholen uit het PRIMA-cohortonderzoek met een vve-aanbod blijkt dat deelname aan een vve-programma geen effect heeft op de leerprestaties van de leerlingen. Bij de analyse zijn de volgende invoerings- en programmakenmerken meegenomen:

- gebruik vve-programma;
- samenwerking met een peuterspeelzaal;
- actieve ouderbetrokkenheid;
- afstemming met groep 3;
- vve-training voor leerkrachten van groep 1 en 2;
- pedagogische benadering (programmagestuurd of ontwikkelingsgericht).

De auteurs stellen dat het ontbreken van een experimentele opzet van het onderzoek beperkingen oplevert om effecten te kunnen vinden. Daarnaast speelt de onvolledigheid en de beperkte bruikbaarheid van de leerlingdata uit het PRIMA-onderwijscohort ook een rol (Nap-Kolhoff *et al*., 2008).

Voor- en vroegschoolse educatie is bedoeld voor kinderen met achterstanden.
Gemeenten bepalen zelf welke kinderen tot de doelgroep behoren. Naast deze definitiekwestie speelt een rol dat steeds meer niet-doelgroepkinderen gebruik maken van het aanbod en niet alleen de kinderen voor wie het programma in eerste instantie bedoeld is. Op steeds meer reguliere kinderdagverblijven worden vve-programma's aangeboden. Uit de Landelijke Monitor Voor- en Vroegschoolse Educatie (Beekhoven *et al*., 2010) blijkt dat bij 60% van de achterstandsscholen met een programma van enkele dagdelen, alle kinderen aan het programma deelnamen. Driessen (2012) stelt dat in het huidige overheidsbeleid ervan uit wordt gegaan dat vve goed is voor alle kinderen, of ze tot de doelgroep behoren of niet. Het is mogelijk dat niet-doelgroep kinderen meer profiteren van het aanbod en meer vooruit gaan dan de doelgroepkinderen. Op deze manier ontstaat er een groot verschil in leerprestaties tussen doelgroep- en niet-doelgroepkinderen. Driessen stelt dat dit een verklaring zou kunnen zijn voor de zwakke of zelfs negatieve effecten die in sommige studies gevonden worden voor de deelname aan de voor- en vroegschoolse educatie.

2.3 Synthese

Internationaal gezien is er consensus over het nut van het investeren in de voor- en vroegschoolse periode (Mutsaers *et al*., 2012; Meij *et al*., 2009). Het effect van vve-programma's in Nederland is, blijkend uit onderzoek, echter niet groot.
Soms wordt alleen gesteld dat het überhaupt deelnemen aan *preschooling* of vve effect sorteert voor de ontwikkeling van jonge kinderen; in andere gevallen worden concrete programma's genoemd. Vaker worden factoren genoemd die een rol (lijken te) spelen bij de geconstateerde effectiviteit.

In 'Vve onder de loep', een studie naar de uitvoering en effectiviteit van voor- en vroegschoolse educatieve programma's (Nap-Kohlhoff *et al*., 2008), wordt op basis van de bestudering van de (internationale) literatuur geconstateerd dat een effectief vve-programma aan een hele lijst kenmerken voldoet. Deze lijst met faal- en succesfactoren wordt in deze synthese uitgebreid met bevindingen uit de voorgaande paragraaf van dit onderzoeksrapport. Ze staan vermeld in schema 2.1.

Schema 2.1 Overzicht faal- en succesfactoren in vve

- adequate pedagogische benadering (zowel leren als sociale ontwikkeling krijgen aandacht)
- gerichtheid op meerdere ontwikkelingsdomeinen
- rijke leeromgeving
- intensiteit van minimaal drie (beter is vier) dagdelen per week (fulltime is niet beslist noodzakelijk)
- deelname gedurende langere tijd
- doorgaande lijn van voor- naar vroegschoolse periode (van peutergroep naar groep 1)
- doorgaande lijn naar groep 3 en later
- kleine groepen en dubbele bezetting (gunstige kind-stafratio)
- goede mix van kinderen met diverse sociaal-culturele achtergronden
- ouderbetrokkenheid (inclusief informatievoorziening en ondersteuning)
- goede, stimulerende leeromgeving thuis (actieve benadering van kind)
- regelmatige evaluatie middels observatie- en/of toetsingsmethoden
- professionaliteit van hooggekwalificeerde uitvoerders (algemeen, programma-specifiek)
- opstellen en evalueren van beleidsplannen
- algemene schoolkenmerken:
  - slagvaardige schoolleiding
  - consensus binnen het team over uitgangspunten en doelen
  - nascholingsplannen voor de uitvoerders
  - hoge verwachtingen ten aanzien van de leerlingen
- thuistaal is Nederlands
- het programma is centrumgericht en wordt op een centrum of op school aangeboden (continuïteit), liefst in combinatie met gezinsgericht programma
- modelprogramma's met bij voorkeur huisbezoek
- kwaliteit van (de uitvoering van) het programma (dit speelt door diverse van bovenstaande factoren heen)

3 Onderzoek interventies tijdens reguliere basisschooltijden

3.1 Inleiding

Niet alleen voorafgaand aan de basisschool, maar ook tijdens de reguliere schooltijden worden interventies gepleegd waarbij gestreefd wordt onderwijsachterstanden te reduceren.
Een voorbeeld hiervan is de aanzet van het project Pilots Taalbeleid Onderwijsachterstanden in 2005 waarbij vanuit het Ministerie van OCW scholen gestimuleerd worden om hun taalbeleid verder te ontwikkelen en er voor te zorgen dat het aantal slecht presterende leerlingen op het gebied van taal en lezen vermindert. Het gemiddelde beheersingsniveau in het Nederlands van tweetalige kinderen blijft achter bij dat van autochtone kinderen. In een gewone klassensituatie is dit verschil moeilijk in te halen. Wanneer in het onderwijs met zaakvakken wordt begonnen, wordt taal een steeds groter struikelblok en worden grote individuele verschillen in taalvaardigheid in het Nederlands duidelijk merkbaar. Om taalachterstanden te bestrijden, kunnen gemeenten en scholen verschillende interventies plegen. Deze komen, voor zover de onderzoeksliteratuur eraan aandacht besteedt, in dit hoofdstuk aan de orde.

3.2 Resultaten

Faubert (2012) maakt met het oog op onderwijsachterstanden een interessant onderscheid tussen het schoolsysteem en de leerling. Als een leerling niet kan meekomen, zou dat volgens hem meer gerelateerd zijn aan het schoolsysteem dan aan de (on)mogelijkheden (cognitief of anderszins) van de leerling. De term '*school failure*' krijgt daarmee een andere dimensie. Om gelijkheid tussen leerlingen te bevorderen, hoeft er in Fauberts visie dan niet meer gekeken te worden naar achtergrondkenmerken van leerlingen zoals geslacht, sociaaleconomische status, etniciteit of handicaps/stoornissen, maar moet het schoolsysteem gericht zijn op onderwijsbehoeften en daarop aansluitende, gedifferentieerde lespakketten aanbieden.

Vanuit deze visie verricht hij een literatuurreview, waaruit blijkt dat *school failure* overwonnen kan worden door interventies op het gebied van:

- pedagogiek (de onderstaande drie punten hebben in hun onderlinge samenhang een vrij sterk effect);
  - instructie (directe instructie versus studentgerichte instructie);
    - herhalen van de lesstof (*space learning over time*);
    - aanbieden van een uitgewerkt probleem, gevolgd door een gerelateerde onopgeloste kwestie;
    - aanpassen van informatie aan het referentiekader van de individuele leerling (*culturally responsive instruction*);
    - met aanschouwelijk onderwijs abstracte begrippen, zoals breuken, verduidelijken;
    - stellen van inzichtvragen in plaats van slechts feitenkennis (*deep questions*);
    - leerlingen in kleine groepjes laten werken, maar die moeten dan niet op basis van vaardigheden samengesteld zijn (negatieve groepssamenstelling (*compositional*) effecten);
  - toetsing (assessment);
    - toetsen met daarop volgend gerichte feedback (nabespreking);
    - kijken wat je al/nog weet, dat checken aan de juiste antwoordenlijst en dan je studietijd focussen (*formative use of summative assessments*);
    - evaluatie of leerplan op schema ligt;
    - vaststellen wat je als leraar eigenlijk wilt weten (*construct definition and interpretation*);

    - curriculum (doelgericht en geïntegreerd, dus niet in eigenstandige lesblokken);
- leiderschap (bescheiden effect);
    - visionair;
    - goed gecommuniceerde en heldere doelen gericht op prestatieverbetering;
    - goed middelenbeheer in relatie tot doelen;
    - facilitering van professionele ontwikkeling van leraren;
    - bewaking van minimum aantal lesgebonden uren;
- onderwijsondersteuning in school en klas (bescheiden effect);
    - data als managementinformatie;
    - professionalisering;
    - enige vrij te besteden, niet-lesgebonden tijd voor leraren;
    - persoonlijk, empathisch contact leraar-leerling *(personalized interventions*);
- school- en klasomstandigheden (bescheiden effect);
    - schoolgrootteverkleining;
    - klassenverkleining;
    - fysieke voorwaarden (bijvoorbeeld temperatuur, licht, akoestiek) (Faubert, 2012).

Muijs *et al*. (2004) stelden een *narrative* review op en betrekken daarin allerlei soorten onderzoek, waaronder kwalitatieve onderzoeksdesigns. Men richt zich daarbij op sociaaleconomisch achtergebleven (*disadvantaged*) scholen. Ze stellen dat leraren op achtergestelde scholen harder moeten werken en meer toegewijd moeten zijn dan elders. Ze moeten dit ook *blijven*, omdat succes hier kortdurend en fragiel is. Blijvende toewijding van leraren aan hun doelgroep is een voorwaarde voor het vasthouden van verbetering op deze scholen. Het is voor deze scholen moeilijker om kwalitatief goed personeel te krijgen en dat ook nog eens te behouden. De zwakke maatschappelijke omstandigheden van de buurt waarin zulke scholen veelal staan (werkloosheid, psychische en lichamelijke ongezondheid, *braindrain,* matige leerprestaties, negatief gedrag van leerlingen), kunnen tot een groot verloop onder leraren leiden (Muijs *et al*., 2004).
Uit hun review komt naar voren dat een aantal strategieën effectief gebleken is bij achterstandsscholen:

- duidelijke focus op een beperkt aantal doelen met een korte termijn, op instructiestrategieën;
- gerichtheid op leerprestaties (*academic orientation*);
- aanbieden van meer structuur aan leerlingen en positieve bekrachtiging door leraren (ook niet-taakgebonden complimenten, vanwege het lagere zelfbeeld van deze leerlingen);
- aanbieden van curriculum in kleinere eenheden, maar wel geïntegreerd (vak- en leerjaar-overstijgend) en met direct daarop volgend een toets (*rapid feedback*);
- meer instructie geven;
- ermee rekening houden dat leerlingen met een onderwijsachterstand veelal beter reageren op extrinsieke beloning;
- verbinden van leren aan ervaringen in het dagelijks leven; praktijkgericht onderwijs;
- hanteren van een consistente en coherente leeraanpak gedurende de hele schoolcarrière;
- meer nadruk leggen op basale vaardigheden als taal en rekenen en minder op verbreding van curriculum, maar toch ook:
- wel een beroep blijven doen op de capaciteiten van leerlingen, hen uitdagingen bieden, hen niet onderwaarderen; hierbij moeten realistische doelen gesteld en positieve feedback gegeven worden;
- continue professionalisering door leraren, gericht op lesgeven (zie verderop);
- hanteren van mix van traditionele en constructivistische lesmethoden;
- gerichtheid op kunstzinnige uitingen en aandacht voor musische vorming;
- ordelijke omgeving;
- duidelijke regels (bij opstellen daarvan leerlingen betrekken).

Al deze punten brengen Muijs *et al*. (2004) samen onder het kopje 'focus op leren en onderwijzen'. Ze vervolgen hun lijst van – in bepaalde settingen – effectief gebleken interventies met een stel aspecten die onder 'leiderschap' te scharen zijn:

- effectieve schoolleider; diens effectiviteit hangt niet zozeer samen met een specifieke leiderschapsstijl, maar omvat eerder gedeelde besluitvorming en collegialiteit, is eerder democratisch dan visionair; verder zorgt een effectieve schoolleider meer voor het proces, laat de (kwalitatieve) inhoud van interventies over aan de leraren en heeft geen managersattitude, maar is eerder initiator; kortom: *collaborative leadership*;
- stimuleren dat de gemeenschap inbreng heeft in schoolverbetering.

Een constructieve schoolcultuur is ook een belangrijk element, blijkt uit de literatuur. Dit wordt door Muijs *et al*. (2004) als volgt ingevuld:
- open communicatie (onder meer over het verwachtingspatroon van de schoolleider);
- samenwerking en vertrouwen;
- gericht op een hoger doel;
- gericht op onderwijs in plaats van administratie en beheer;
- gering verloop onder leraren;
- perceptie van leerkrachten op de effectiviteit van een interventie (eerdere positieve ervaringen werken enthousiasmerend);
- familiegevoel op school.

Effectieve achterstandscholen creëerden vaker een informatierijke omgeving. Hieronder wordt door Muijs *et al*. (2004) het volgende begrepen:
- datagestuurd onderwijs (kwalitatief en kwantitatief onderzoek) gericht op verbetering van schoolprestaties;
- niet alleen toetsen voor leerlingen, maar ook tevredenheidsonderzoek onder leerlingen en personeel.

Een andere groep kenmerken die achterstandscholen helpt op effectieve wijze te verbeteren, betreft het creëren van een *learning community*, waarbij onder meer gedacht kan worden aan continue professionalisering van leraren. Elementen hierbij zijn:
- blijvend leren, *reflective dialogue* en kleine lerarenteams;
- continue professionalisering door leraren, niet gericht op hun persoonlijke doelen maar juist wel praktijkgericht, aanschouwelijk, met theoretisch element en met coaching en feedback;
- met het oog op professionaliseringsactiviteiten tijd vrijstellen en kennisnetwerk onder scholen opzetten (Muijs *et al*., 2004).

Tenslotte komen uit de literatuurreview de volgende aspecten naar voren:
- ouderbetrokkenheid (maak lesstof duidelijk en laat ouders hun kinderen helpen, taalles voor ouders, maatschappelijk werkers);
- externe support (netwerken en *monitoring*); soms werken inspecties positief;
- adequate mate van financiering (voorwaarde voor effectiviteit: goed management) (Muijs *et al*., 2004).

Bovenstaande punten zijn in de literatuur aangewezen als effectieve elementen aan de hand waarvan achterstandsscholen beter gaan presteren. Zonder bestudering van alle onderliggende artikelen en onderzoeksrapporten, kan geen goed beeld gevormd worden van de robuustheid van elk van de geïnventariseerde effectieve aspecten. In de hier gepresenteerde opsomming geldt elke vermelde factor als even zwaar als de overige, los van door hoeveel onderzoeken deze onderbouwd is en aan de hand van welk onderzoeksdesign men tot de geconstateerde effectiviteit besloot (in hoofdstuk 5 wordt hierop veel dieper ingegaan).

Muijs *et al*. (2004) gaan nog verder door enkele voorwaarden op te sommen vanuit de onderzoeksliteratuur die helpen schoolverbetering te consolideren:
- leraren moeten methoden als effectief zien;
- schoolleiders moeten de verandering managen en ondersteunen;
- er moet een cultuur van continue professionalisering zijn;
- er wordt actief goed gekwalificeerd personeel geworven;

- de succesvolheid wordt continu in de gaten gehouden;
- de school gedraagt zich als een lerende organisatie;
- nieuwe inzichten worden in de praktijk ingevoerd;
- er is ondersteunend gemeentelijk beleid;
- er is stabiel leiderschap (visie).

Een interessant onderzoek wordt verricht en beschreven door Van Ewijk & Sleegers (2010). Zij voerden een meta-analyse uit op dertien studies naar de invloed van de samenstelling van klassen op leerprestaties van leerlingen. Zij concluderen dat in etnisch gemengde klassen waarin één etnische groepering met een groot aandeel leerlingen vertegenwoordigd is, schoolresultaten dalen voor leerlingen behorend tot die etnische groep. De leerprestaties van eventuele andere etnische minderheden noch van leerlingen van autochtone afkomst worden hierdoor beïnvloed. Zij noemen dit het *compositional effect* en constateren dat dit slechts een klein effect is. Waarom het toch van betekenis kan worden, is door een mogelijk cumulatie-effect, waarbij door de dominante etnische minderheid per schooljaar steeds een klein minnetje gescoord wordt, optellend tot een forse min aan het einde van de schoolcarrière.
Tenslotte nog een algemene bevinding. In de verzameling van 800 meta-analysen van Hattie (2009) over wat werkt in het onderwijs, wordt gezien dat er vaak maar kleine effecten worden gevonden. Sterke effecten worden gevonden voor moeilijk meetbare zaken zoals het feit dat leerkrachten hoge verwachtingen hebben van leerlingen en hen kunnen uitdagen om na te denken en problemen op te lossen. Als deze aanmoediging ook van de kant van de ouders komt, dan krijgen de kinderen optimale kansen voor het schoolse leren. Grootste 'thuiseffect' is gelegen in de hoge aspiraties en verwachtingen van ouders. Hoe hoger de verwachtingen, des te beter de prestaties. Dit levert een sterk effect op met schoolprestaties (d=.80). Als kinderen beginnen op school, heeft 98% van de ouders hoge verwachtingen over de belangrijkheid van onderwijs. Als kinderen de basisschool verlaten, hebben veel ouders hun verwachtingen naar beneden bijgesteld.
Sociaaleconomische status (*SES*) speelt bij verwachtingen een rol. Bij een lage *SES* is er een barrière tussen school en thuis. De 'schooltaal' is onbekend bij lage *SES*-ouders. In de meta-analysen van Hattie (2009) zijn effecten bekend van projecten waarbij ouders worden geholpen om de 'schooltaal' te verstaan en hun kinderen te ondersteunen bij schoolwerk (maar dan niet op een controlerende manier) (Clinton *et al*., 2007; Jeynes, 2007). Ouders zouden moeten worden onderwezen in 'schooltaal', zodat de verwachtingen over het kind hoog zijn zowel thuis als op school. De psychosociale omgeving en intellectuele stimulering van de kinderen thuis is een belangrijke voorspeller voor schoolsucces (effectgrootte d=.57) met als ingrediënten de betrokkenheid van de moeder, goed speelgoed en de mogelijkheid voor kinderen om gevarieerde ervaringen op te kunnen doen.

3.3 Synthese

Er zijn enige parallellen te zien tussen de review van Muijs *et al*. (2004) en die van Faubert (2012). Zo bevelen beiden de gerichtheid op instructiemethoden aan, zien de focus op leerprestaties als effectief en hebben het over een positieve relatie tussen leraar en leerling. Verder vinden ze elkaar als het gaat om kleine, geïntegreerde brokken lesstof met daarop volgend een toets, waarbij de lesstof praktijkgericht moet zijn. Ook over de professionalisering van leraren en het nut voor leraren voor een paar uren vrijgesteld te zijn van onderwijstaken om bijvoorbeeld aan die professionalisering toe te kunnen komen, zijn de auteurs het eens. Tenslotte noemen beiden nog ruimtelijke omgevingsfactoren en een goed middelenbeheer door de schoolleider.
Over het al dan niet visionair zijn van de effectieve schoolleider verschillen de auteurs echter van mening. Verder onderkennen beiden het belang van datagebruik, maar gaan daarmee elk een andere kant op: Faubert (2012) gebruikt data als managementinformatie, terwijl Muijs *et al*. (2004) die inzetten ten behoeve van verbetering van leerprestaties.
In situaties waarin multiculturele klassen voorkomen, moet rekening gehouden worden met een *compositional effect*. Dit verschijnsel treedt echter ook op bij het formeren van groepen leerlingen op basis van hun vaardigheden.

Net als gevonden werd bij *preschooling* en vve, draagt het aan de effectiviteit van interventies ter bestrijding van onderwijsachterstanden bij als leerkrachten en ouders hoge verwachtingen koesteren van de aan hen toevertrouwde kinderen.
In schema 3.1 wordt een overzicht geboden van de gevonden factoren, die effectief blijken te zijn bij het bestrijden van onderwijsachterstanden. Dit zijn factoren die behoren tot goed onderwijs in het algemeen, maar juist in achterstandssituaties kunnen deze factoren het verschil maken voor een leerling.

Schema 3.1 – Overzicht factoren interventies tijdens reguliere basisschooltijden

- pedagogiek: focus op leren en onderwijzen [grootste effect]
  - instructie gericht op leerprestaties; professionalisering (*learning community*), structuur, beloning, uitdagingen bieden, hoge verwachtingen hebben van leerlingen
  - toetsing
  - praktijkgericht, geïntegreerd curriculum (waaronder kunst-onderwijs)
- schoolleiderschap (procesgericht, *collaborative* leiderschap*)*
- constructieve schoolcultuur (open communicatie, samenwerking, gering verloop onder leraren, perceptie van leerkrachten op de effectiviteit van een inter-ventie, familie-gevoel op school, empathisch contact leraar-leerling)
- ouderbetrokkenheid ('schooltaal' leren verstaan en kinderen ondersteunen bij schoolwerk; psychosociale omgeving en intellectuele stimulering thuis)
- informatierijke omgeving (datagestuurd onderwijs, toetsen voor leerlingen, tevredenheidsonderzoek onder leerlingen en personeel)
- school- en klasomstandigheden (schaalverkleining, *compositional effect,* fysieke voorwaarden)
- adequate mate van financiering
- externe support (netwerken en monitoring)

4 Onderzoek interventies na reguliere basisschooltijden: afterschool programs

4.1 Inleiding

Dit hoofdstuk gaat over interventies die gepleegd worden buiten de reguliere schooltijd om. De organisatievormen zijn velerlei. Ze hebben gemeen dat de verlenging van de onderwijstijd een verlenging van de effectieve leertijd beoogt, waarin de extra tijd wordt benut voor het actief tegengaan van onderpresteren bij leerlingen, in het bijzonder achterstandsleerlingen (Appelhof, 2009). Andere redenen voor het aanbieden van verlengde leertijd zijn het voorkomen van terugval in taal- en vooral rekenprestaties (*achievement gap*) gedurende de lange zomervakanties in met name Amerika en ook het voorzien in opvang als ouders aan het werk zijn ter preventie van negatief gedrag (Van den Bulk & Algera, 2012).
Zo kan een *afterschool program* eruit bestaan dat volgend op de reguliere schooldag aan leerlingen nog een aantal uren wordt aangeboden. De inhoud van die uren verschilt per programma; sommige zijn eerder op te vatten als bijles of huiswerkbegeleiding, andere besteden meer aandacht aan niet-cognitieve elementen, zoals kunstzinnige of culturele uitingen. In de Engelstalige onderzoek literatuur wordt dit ook wel *extended schoolday* genoemd.
De extra uren, die aan bepaalde leerlingen worden aangeboden, kunnen ook geconcentreerd worden in een aantal extra lesweken. Hiervan is sprake bij zomerscholen (*summer schools*). Dit is een vorm van *extended schoolyear*.
Een nog verderstrekkende vorm is een heel jaar bijspijkerles, in de vorm van de Nederlandse schakelklassen. Hoewel schakeluren ook als *extended schooldays* gegeven kunnen worden (dus naast het gebruikelijke curriculum), is een heel tussenjaar, op verschillende momenten in te lassen tijdens de basisschoolloopbaan, gebruikelijker. Als dit extra leerjaar aan groep 8 vastgeplakt wordt, dan wordt de schakelklas een kopklas genoemd. Doel is om aan leerlingen met een taalachterstand extra taallessen aan te bieden, waardoor ze na verloop van tijd goed kunnen meekomen met de reguliere klas dan wel (in het geval van een kopklas) kunnen doorstromen naar een hogere vorm van voortgezet onderwijs.
Twee voor Nederland vrij nieuwe, maar in het buitenland reeds beproefde programma's zijn het Britse *Playing for Success* en de *Children's University*. Het eerstgenoemde programma is erop gericht het zelfvertrouwen van onderpresterende leerlingen aan te wakkeren, waarbij sportidolen een belangrijke plaats innemen (Appelhof, 2009). Het tweede programma biedt leerlingen innovatieve leeromgevingen aan die geworteld zijn in de samenleving (MacBeath, 2011). Beide programma's zijn op te vatten als aanvullingen op het curriculum buiten de reguliere schooltijden.

4.2 Resultaten

Clark *et al*. (2005) onderzochten hoe Afro-Amerikaanse mannen zich ontwikkelden, nadat ze een *afterschool* programma doorlopen hadden. Ze definiëren een achttal '*selves*', waarvan mentale, cognitieve vaardigheden er één zijn. Zij beschouwen elf studies, die toetsresultaten en (voor dit literatuuronderzoek minder interessant) spijbelgedrag als afhankelijke variabele hebben en deelname aan structurele *afterschool* activiteiten als de onafhankelijke. In hun literatuurreview vinden zij dat het participeren in schoolgerelateerde extracurriculaire activiteiten leidt tot betere schoolprestaties dan wanneer die buitenschoolse activiteiten niet schoolgerelateerd zijn. Verder noemen zij nog enkele voorwaarden waaronder ouders en kinderen gunstiger percepties ontwikkelen. Omdat percepties geen 'bewijzen' van effectiviteit zijn, wordt hierop niet verder ingegaan.

In hun systematische review nemen Patall *et al*. (2010) de verlengde schooldag respectievelijk het verlengde schooljaar onder de loep. Zij doen dit met behulp van onderzoeken die zich richten op leerlingen van 4 tot 18 jaar en dus eigenlijk de hele schoolcarrière omvat. Omdat de verwachtingen in de Verenigde Staten aangaande verlengde leertijd hooggespannen zijn in relatie tot betere

leerprestaties, gaan zij na hoe de relatie tussen die beide nu precies ligt. Zij sommen een hele lijst met voor- en nadelen op, zoals die door betrokkenen (leerlingen, leerkrachten, ouders en de maatschappij als geheel) kunnen worden ervaren. Hoewel ze constateren dat het verrichte onderzoek veelal correlationeel van aard is, waardoor er geen causale verbanden gelegd kunnen worden, beschrijven zij een artikel (Silva, 2007) waarin conclusies over verlengde leertijd getrokken worden. Daar wordt gesteld dat verlengde leertijd vooral nuttig kan zijn voor *at risk* leerlingen, voor leerlingen van wie de ouders een laag inkomen hebben of voor gehandicapte leerlingen. Randvoorwaardelijk blijkt te zijn dat scholen een effectief curriculum hebben, goede instructie geven en een adequaat klassenmanagement hanteren (Patall *et al*., 2010).
De auteurs baseren hun review op vijftien onderzoeken die over de relatie tussen verlengde leertijd en schoolprestaties gaan. Hoewel ze kort de gebruikte designs beschrijven (descriptief, correlationeel, cohortstudies en (quasi-)experimenten), spreken zij zich niet uit voor de ene of de andere onderzoeksopzet. Toch bevelen ze later aan een sterk design te kiezen, opdat daaruit causale verbanden afgeleid kunnen worden.
Enkele studies lijken in de richting te wijzen dat er voor *sommige* subgroepen (leeftijdsgroepen, etnische minderheden, zwakke leerlingen, lage *SES*) aanwijzingen te vinden zijn dat verlengde leertijd ertoe doet, maar die effecten gelden ook weer niet voor alle uitkomstmaten. Het enige onderzoek met een experimenteel design kent de zwakte, dat de interventie (langere schooldag) vertroebeld wordt door het volgen van meer vakken in plaats van langer dezelfde vakken te volgen. Een verlengd schooljaar lijkt iets beter te werken dan een verlengde schooldag, zeker voor bepaalde *SES*-groepen en *at risk* jongeren. Extra leertijd is ook gunstig voor buitenlandse studenten die naast de betreffende vakken zich ook het Engels moeten eigen maken; het lijkt wat op dyslectische kinderen die wat langer over hun toets mogen doen.
Het is dus allemaal vrij summier, ook al omdat de studies waarin tenminste nog enig effect gevonden wordt, veelal analyseproblemen kennen: de leerling wordt als analyse-eenheid aangehouden hoewel op schoolniveau de interventie toegewezen wordt. Bovendien zijn de experimentgroepen maar van beperkte omvang. De enige studie die van deze tekortkomingen geen last heeft, vindt echter geen significante effecten.

Lauer *et al*. (2006) verrichtten een meta-analyse van de effecten van *out-of-school* programma's voor *at risk* leerlingen gedurende hun schoolcarrière (4-18 jaar). Zij vinden kleine, positieve effecten, waarbij het doorgaans niet zoveel uitmaakt of de interventie nu *afterschool* of gedurende de zomervakantie plaatsvindt. Zij stellen op basis van hun studie dat enkele omstandigheden, waaronder *out-of-school time* (*ost*) plaatsvindt, belangrijk zijn, te weten: de participatiegraad (al dan niet vaak de *ost*-activiteiten bezoeken) en prestatiegraad (laagpresteerders hebben er meer aan dan cognitief zeer succesvolle leerlingen). Verder lijkt het raadzaam de *ost*-activiteiten aan te passen al naar gelang de leeftijd. Zo lijkt het belangrijker voor oudere leerlingen om meer recreatieactiviteiten aangeboden te krijgen dan dat het voor jongere kinderen het geval is. De meningen van onderzoekers zijn volgens Lauer *et al*. (2006) verdeeld aangaande de mate waarin de *ost*-activiteiten zouden moeten aansluiten op schoolse (cognitieve) vaardigheden. Sommigen vinden van wel, anderen leggen meer nadruk op andere bezigheden. Ook Lauer *et al*. (2006) lopen in hun meta-analyse tegen de zwakte van veel gebruikte designs aan in hun poging stevige conclusies te trekken.
Als alle resultaten zo verantwoord mogelijk bij elkaar worden gevoegd, dan kan met betrekking tot verbeteringen in de leesvaardigheid geconcludeerd worden dat het niet zoveel uitmaakt of leerlingen *afterschool*-activiteiten bezoeken of een *summer school* (vergelijk het onderscheid tussen *extented schoolday* en *extended schoolyear*). Ook de focus van het programma (schoolgerelateerd ('*academic*') of sociaal) doet er niet significant toe. Programma's met een middellange duur (44-210 uren) blijken effectiever dan kortere of nog langere programma's. Leerlingen in wat in Nederland zo ongeveer overeenkomt met de middenbouw (*grade* 3-5) scoren geen significante *effectsize*, leerlingen in andere clusters (onderbouw, bovenbouw, voortgezet onderwijs) hebben gemiddeld genomen wel baat bij *ost*. Vooral het één-op-één met leerlingen werken boekt succes, in tegenstelling tot grotere of kleinere groepen. De kwaliteit van het onderwijs blijkt ook een moderator: in geval van medium en *high quality* onderwijs werkt ost beter dan bij onderwijs van lage kwaliteit.
Op zoek naar interactie-effecten tussen de verschillende omstandigheden waaronder *ost* aangeboden wordt, vinden Lauer *et al*. (2006) dat bijvoorbeeld één-op-één tutoring alleen in *afterschool*

programma's voorkomt en nooit op de zomerschool. Als belangrijke aspecten van *ost*-programma's worden aangemerkt de participatiegraad, de kwaliteit van de supervisor en of het curriculum doortimmerd is.
Voor rekenen gelden gedeeltelijk dezelfde resultaten: de vorm van *ost* (*summerschool* of *afterschool*) doet er niet toe en het werkt beter in settings met kwalitatief goed onderwijs. De verschillen zijn gelegen in dat *ost* juist niet voor de onderbouw nuttig is, dat de duur minimaal 46 uur moet zijn (geen bovengrens) en dat het juist niet één-op-één moet plaatsvinden maar in groepsverband, groot of klein. Wat betreft rekenen helpen ook programma's die niet uitsluitend cognitief van aard zijn, maar zich ook op de sociale aspecten van de leerling richten.
Al met al concluderen Lauer *et al*. (2006) dat er zeker effecten gevonden kunnen worden, zij het dat daarin niet zozeer een consequente lijn te ontwaren is en dat de onderliggende designs veelal zwak zijn. Kassenberg (2010) verricht een literatuurreview en duidt haar bronnen met ID-nummers aan (zonder overigens een bibliografie op te nemen). Uit haar beschrijving (ID-123) wordt echter duidelijk dat zij ook aan het onderzoek van Lauer *et al*. (2006) refereert.

Kassenberg (2010) beschrijft nog een ander onderzoek (ID-122) onder lage *SES*-kinderen. Hieruit blijkt dat kwalitatief hoogstaande, naschoolse programma's, gericht op ondersteuning bij schoolprestaties, ontspanning, kunst en verrijkingsactiviteiten effectief zijn bij de doelgroep. Er is sprake van een sterke relatie tussen staf en jongeren, grote betrokkenheid bij de programma's door jongeren zelf en een gestructureerd aanbod van de activiteiten. Mede door dit alles worden – naast sociale winst – hogere rekenscores gemeten bij leerlingen die regelmatig deelnemen aan het naschoolse programma. Kassenberg (2010) benoemt mede op grond van dit onderzoek als voorwaarden voor een effectieve terugdringing van onderwijsachterstanden:
- individuele aandacht voor leerlingen tijdens de naschoolse activiteiten;
- betrokkenheid van ouders;
- intensieve deelname.

Woodland (2008) beschrijft een situatie waarin jonge Afro-Amerikanen, die naschools zonder toezicht verblijven, sneller tot ongewenst gedrag vervallen. *Afterschool* programma's moeten hier soelaas bieden. In zijn verhalende literatuurreview beschrijft hij kernelementen van effectieve *afterschool* programma's:
- relaties tussen volwassenen en kind (door nauw contact meer zorg en ondersteuning);
- flexibiliteit (aanpassen bij de behoeften van de groep, maar toch altijd gericht zijn op sociale waarden en hoge verwachtingen koesteren);
- training van medewerkers (kunnen beter omgaan met doelgroep en stralen mee gezag uit);
- veilige ruimte (fysiek, psychisch en sociaal);
- culturele component (aandacht voor culturele waarden van etnische groepen);
- betrokkenheid van familie (rekening houden met belangen van het gezin);
- verrijkend curriculum (uitdagende ontwikkelingsmogelijkheden bieden, ook op het vlak van kunst, muziek en dans);
- één-op-één hulp bij het leren (bijvoorbeeld huiswerkbegeleiding; hiermee wordt een koppeling gelegd tussen *afterschool*-activiteiten en de reguliere schooldag);
- grondige evaluatie (waardoor het programma steeds verbetert en betrokkenen tevreden blijven).

Tenslotte noemt Woodland (2008) als een cruciale determinant van een succesvol *afterschool* programma een onvoorwaardelijk geloof in de mogelijkheden van leraren en/of *afterschool* programmabegeleiders, ingebed in het besef van culturele relevantie.

Terzian *et al*. (2009) richten hun aandacht op zomerscholen voor *low-SES* leerlingen. Ze stellen dat door de lange zomervakantie (in Amerika) er een terugval te zien is in taal- en rekenprestaties van leerlingen. Deze groep leerlingen heeft thuis minder toegang tot boeken en bij hen kan minder het gebrek aan structuur, die school biedt, thuis opgevangen worden. In de Verenigde Staten zijn vervoer naar en kosten van *summerschools* problematisch; de doelgroep is dan ook lastig te bedienen.
Effectieve programma's hebben volgens de auteurs in hun beleidsbrief gemeen dat leren leuk gemaakt wordt, lesstof toegepast wordt op het werkelijke leven en doedingen geïntegreerd worden in het lesprogramma. Verder sluiten zulke programma's aan op het reguliere curriculum, zijn de groepen klein

(10–15 leerlingen, met een *adult/child*-ratio van 1:5 vanwege de persoonlijke aandacht) en is de kwaliteit van de professionals hoog. Tenslotte betrekken dergelijke, effectieve programma's ouders en de gemeenschap erbij en zijn ze doelgericht (Terzian *et al*., 2009).

Mulder *et al*. (2012) tonen aan dat schakelklassen een waardevol en effectief instrument zijn om in te zetten ter bestrijding van onderwijsachterstanden. In een quasi-experiment volgen zij vier jaar lang leerlingen die op verschillende momenten in hun basisschoolloopbaan een schakelklas doorliepen of in een kopklas zaten (schakelklas na groep 8).
Dit onderzoek biedt handvatten voor de bespreking van geschikte onderzoeksdesigns om de effectiviteit van interventies vast te stellen. In hoofdstuk 5 zal het onderzoek daarom in extenso besproken worden en wordt de aandacht hier beperkt tot enkele conclusies.
De eerste vervolgmeting, zo'n twee jaar na de nulmeting, wijst uit dat de voormalige schakelklasleerlingen goed kunnen meekomen op het terrein van begrijpend en technisch lezen. Op woordenschat en rekenen is hun achterstand toch weer wat opgelopen. In het tweede vervolgonderzoek, zo'n vier jaar na deelname aan de schakelklas, is die dip weer verdwenen. Kopklassers doen het in beide vervolgonderzoeken beter dan (qua sociaaletnische herkomst en Cito-eindtoetsscore) vergelijkbare leerlingen die niet in een kopklas gezeten hebben. Zo is het onderwijsniveau waarop ze het voortgezet onderwijs instromen hoger en kunnen ze het niveau aldaar beter aan.

De *Children's University* in Engeland is een nieuwe, projectmatige samenwerking van universiteiten, bedrijfsleven en/of non-profit organisaties, waarbij kinderen in achterstandswijken buiten de schoolsetting in een educatieve omgeving nieuwe ervaringen kunnen opdoen en samenwerkend kunnen leren. Actieve deelname van de kinderen wordt beloond met een diploma dat op een officieel moment wordt uitgereikt. MacBeath (2011) bestudeerde de toetsresultaten van 3.000 leerlingen die hebben deelgenomen aan de *Children's University* en concludeert dat deelnemende leerlingen meer participeren in het onderwijs, minder uitvallen en dat hun schoolprestaties significant verbeteren ten opzichte van leerlingen op dezelfde scholen die niet deelnemen aan de *Children's University*. Langetermijneffecten zijn nog niet bekend.

Omdat een veelgemaakte opmerking is, dat het onderwijssysteem van bijvoorbeeld Nederland uniek is, waardoor het lering trekken uit buitenlands onderzoek bij voorbaat beperkt is, is een studie (deels) binnen de Nederlandse context extra interessant. Het betreft een recent rapport van Appelhof (2009), naar de effecten van verlengde onderwijstijd. De onderzoekers constateren dat in onderzoeksverslagen veelal niet vermeld wordt of de onderzoekpopulatie een onderwijsachterstand heeft opgelopen dan wel dat zij een niet verbijzonderd onderdeel uitmaken van de onderzochte groep. Ook stellen zij dat aan bepaalde randvoorwaarden voldaan moet zijn, wil er van effectieve bestrijding van onderwijs-achterstanden gesproken kunnen worden. Deze zijn samen te vatten als 'effectief onderwijs' en omvatten:

- hoe verwachtingen stellen aan leerlingen;
- positief belonen ter bevordering van het zelfvertrouwen van leerlingen;
- *time on task*;
- goede instructiekwaliteit;
- continuïteit in het leren;
- motiverende context met verrijkingsactiviteiten.

Als aan deze voorwaarden voldaan is zowel in de reguliere schooltijd als in de naschoolse tijd, dan worden de effecten versterkt.

De implementatie van verlengde onderwijstijd wordt gefaciliteerd indien:

- er stevig leiderschap is (*commitment*);
- leerkrachten zich gedreven inzetten;
- onderwijs datagestuurd is;
- de omgeving (ouders, onderwijsbonden) betrokken zijn;
- de focus op de academische doelen ligt en op toegevoegde verrijking.

Allerlei aanbevelingen van Oberon geven suggesties voor hoe aan deze voorwaarden te voldoen.

Als laatste bron wordt een recent rapport van de Onderwijsraad over Uitgebreid Onderwijs opgevoerd (Onderwijsraad, 2011). Hierin wordt verklaard waarom uitgebreid onderwijs met name effect heeft bij achterstandsleerlingen. Er wordt gesteld dat deze leerlingen in het naschoolse programma de prikkels krijgen aangeboden die ze thuis missen om zich te ontwikkelen. Kinderen in achterstandswijken hebben minder toegang tot vrijetijdsvoorzieningen dan anderen en ook nemen zij te weinig deel aan het sociale leven. Naschoolse programma's kunnen voor deze kinderen een effectief en aantrekkelijk alternatief zijn.
Net als in dit onderzoeksverslag gedaan wordt, concludeert ook de Onderwijsraad (2011) dat voor effectieve programma's volgens de Amerikaanse literatuur niet één doorslaggevende succesfactor aan te wijzen is. Voorwaarden voor effectiviteit kunnen wel opgesomd worden:

- intensieve deelname gedurende een langere tijd (minstens meerdere uren per week gedurende minstens een jaar; het optimum is onbekend);
- kwaliteit van het programma (met als kenmerken van een kwalitatief goed aanbod:
  - een inhoudelijk degelijk programma;
  - goed en specifiek opgeleide leraren of activiteitenbegeleiders;
  - goed klassenmanagement;
  - flexibele programma's (niet te strak en te schools, maatwerk);
  - diverse vormen van hulp en begeleiding;
  - niet te grote groepen zodat voldoende individuele aandacht mogelijk is;
  - goede *monitoring* van voortgang;
  - bij samenwerking met andere instanties: goede afstemming, ook tussen regulier en uitgebreid onderwijs;
- een mix van interventies (geen eenzijdige programma's, zoals alléén gericht op sport of leren);
- motivatie van deelnemers (leerlingen – en soms ook hun ouders – moeten continu door leerkrachten geënthousiasmeerd worden) (Onderwijsraad, 2011).

4.3 Synthese

De conclusie van dit hoofdstuk luidt dat er aanwijzingen zijn dat *afterschool* programma's – in welke vorm dan ook – kunnen werken, maar dat niet eenduidig is vast te stellen welke elementen nu het meeste effect hebben, onder welke omstandigheden, voor wie en met het oog op welke effectmaat. De onderzoeksliteratuur komt veelal niet verder dan het benoemen van enkele voorwaarden voor effectiviteit. Niet alleen de variëteit van onderzoeksvragen (en dus andere doelgroepen, afhankelijke en onafhankelijke variabelen), maar ook de zwakte van veel van de gebruikte onderzoeksopzetten is hieraan debet.
Afterschool programs blijken in meer of mindere mate geschikt voor *at risk* kinderen, bijvoorbeeld vanwege lage *SES*, etniciteit of handicap. Er zijn aanwijzingen vanuit de literatuur dat de vorm waarin de extra uren worden aangeboden, ondergeschikt lijkt. Andere studies lijken een lichte voorkeur voor een *extended schoolyear* ten opzichte van een *extended schoolday* te hebben. Schakelklassen en de *Children's University* lijken een effectieve methode om onderwijsachterstanden tegen te gaan. In hun *narrative* review concluderen Mutsaers *et al*. (2012) dat de resultaten van de verlengde schooldag, schakelklassen, naschoolse opvang en vakantiekampen bescheiden zijn.

Uit de onderzoeksliteratuur wordt duidelijk dat een aantal factoren als effectief geldt in het uitvoeren van interventies ter bestrijding van onderwijsachterstanden. Deze staan opgesomd in schema 4.1.

Schema 4.1 – Overzicht factoren interventies na basisschooltijden: afterschool programs

- schoolgerelateerde extra-curriculaire activiteiten, hoewel anderen juist niet-schoolgerelateerde (verrijkings)activiteiten aanbevelen
- effectief curriculum
- kwaliteit van onderwijs
- goede instructie in de les
- hooggekwalificeerde medewerkers
- adequaat klassenmanagement
- participatiegraad (al dan niet vaak de activiteiten bezoeken)
- prestatiegraad (laagpresteerders hebben er meer aan dan cognitief zeer succesvolle leerlingen)
- middellange duur (44-210 uren)
- afgestemd op leeftijdsgroep
- mix van interventies (geen eenzijdigheid)
- één-op-één tutoring
- emotionele band tussen staf en leerlingen
- ouderbetrokkenheid
- veilige ruimte (fysiek en emotioneel)
- positief belonen van leerlingen (zelfvertrouwen bevorderen)

5 Onderzoeksdesigns

5.1 Inleiding

De voorgaande drie hoofdstukken belichten wat effectief is in het bestrijden van onderwijsachterstanden, elk vanuit hun eigen perspectief. De onderzoeksresultaten blijken divers te zijn: onderzoekspopulaties en interventies zijn zeer uiteenlopend en dientengevolge ook de bevindingen. Niet in de laatste plaats variëren de onderzoeksdesigns: niet alleen in het 'soort' onderzoek (binnen het hele scala van kwalitatief tot kwantitatief onderzoek), maar ook in de aard ervan (eigenstandig onderzoek, *narrative* reviews of meta-analysen). Als er iets duidelijk wordt na het doornemen van de internationale onderzoeksliteratuur, is het wel dat er veel kritiek op onderzoeksontwerpen wordt gehoord. Veel onderzoek wordt in meer of mindere mate gediskwalificeerd vanwege tekortkomingen in de opzet ervan. Dit schaadt de 'bewijskracht' van onderzoek in hoge mate en daarmee blijft de werkzaamheid van een bepaalde innovatie of een bepaald programma diffuus.

De vierde onderzoeksvraag (§ 1.2) gaat over hoe methodologisch verantwoord onderzoek te doen naar de effectiviteit van interventies ter bestrijding van onderwijsachterstanden in het primair onderwijs.
Veelal geven de discussieparagrafen in wetenschappelijke artikelen wel enig zicht op mogelijke valkuilen. In dit hoofdstuk worden ze opgesomd (§ 5.2) om daarna (in § 5.3) te komen tot overwegingen voor het kiezen voor een bepaalde onderzoeksmethode, waarvan er vervolgens drie worden gepresenteerd in § 5.4. Dan volgt een intermezzo (§ 5.5), verzorgd door SEOR, waarin geïnventariseerd wordt op welke wijze gebruik gemaakt kan worden van reeds verzamelde data. Dit alles leidt in § 5.6 tot een uitwerking van het door de opdrachtgever in een voorbespreking gekozen onderzoeksdesign.

5.2 Resultaten

De waarschuwingen van auteurs voor methodologische beperkingen van de bewijskracht van onderzoeksdesigns laten zich onderverdelen in zes categorieën. Het gaat hierbij om:

- aard van onderzoek;
- onderzoekspopulatie;
- verstorende variabelen;
- interventie;
- uitkomstmaat;
- setting.

Aard van onderzoek
Patall *et al*. (2010) en Faubert (2012) concluderen in hun (systematische) reviews dat veel verricht onderzoek correlationeel van aard is. Dat wil zeggen dat met dergelijke designs aangetoond wordt dat een bepaalde interventie positief (of eventueel negatief) samenhangt met de bestudeerde uitkomstmaat. Er kan echter geen uitspraak gedaan worden over de causaliteit van het verband. Is de (bijvoorbeeld) hogere waarde voor de afhankelijke variabele een rechtstreeks gevolg van de onafhankelijke variabele? Dit kan alleen overtuigend vastgesteld worden aan de hand van een experimentele onderzoeksopzet, waarbij sprake is van het gerandomiseerd toedelen van de interventie aan de onderzoekspersonen, van manipulatie van de interventie en van een gematchte controlegroep. In 25 jaar tijd – de tijdspanne van hun systematische review – is slechts één gerandomiseerde interventietoewijzing aangetroffen (Patall *et al*., 2010).
In het SCP-rapport 'Op zoek naar Bewijs' staat een schema opgenomen op welke wijze allerlei beleidsmaatregelen getoetst zijn. Verklarende evaluaties en monitors (cohortvergelijkingen) voeren de boventoon. Deze geven weliswaar inzicht, maar zijn designtechnisch niet in staat effecten aan te tonen. Hoe verder men vervolgens opschuift op het continuüm van effectmetingen in de richting van het meest

robuuste design, des te dunner de spoeling wordt (Herweijer & Turkenburg, 2012). Bemoedigend is wel dat Patall *et al*. (2010). kunnen constateren dat juist de sterkere designs consistentere en positievere effecten aantonen van *additional schooltime*.
Muijs *et al*. (2004) staan een gecombineerd onderzoeksdesign voor. Zij stellen dat er beter begrip ontstaat als kwantitatief onderzoek aangevuld dan wel verrijkt wordt door etnografische onderzoeksmethoden en interviews. Meer in het algemeen onderschrijven zij de contingentietheorie als onderzoeksmodel, waarbij het onderzoeksdesign zo goed mogelijk aansluit bij de context waarin het onderzoek plaatsvindt (*best fit to situational circumstances*).
Daarnaast levert volgens Clark *et al*. (2005) een mix van dataverzamelingsmethoden goede resultaten op: kwalitatief onderzoek (interviews en observaties), kwantitatief onderzoek (toetsresultaten en vragenlijstonderzoek) en dossieronderzoek.
Ledoux *et al*. (2011) verrichtten een descriptief onderzoek naar de schoolloopbaan van Nederlandse kinderen. Zij stellen dat het summum van bewijskrachtig onderzoek – het zuivere experiment – niet mogelijk is bij beleidsevaluaties, omdat de interventie niet constant gehouden kan worden en de onderlinge samenhang van tal van factoren (waaronder andere beleidsmaatregelen) erg complex is om te onderzoeken (Faubert, 2012). Het SCP stelt in een onderwijsrapport ook al vast dat veel beleidsinterventies met elkaar verweven zijn en dat de doelrealisatie plaatsvindt in een multi-level- en multi-actorcontext (Herweijer & Turkenburg, 2012).
Onomwonden stelt Kassenberg (2010; ID-132) dat veel conclusies door auteurs helemaal niet getrokken zouden mogen worden, omdat de zwakte van het gebruikte design dat niet verantwoord toelaat. Praktische bezwaren belemmeren veelal een adequaat design; vooral op het vlak van de selectie van de onderzoekspopulatie moeten vaak concessies gedaan worden, met nadelige gevolgen voor de bewijskracht. Als voorbeeld wordt genoemd de toewijzing van alle leerlingen van een hele school aan de experimentgroep in plaats van random toewijzing per leerling. Burger (2010) noemt als voorbeelden van goede onderzoeksdesigns: quasi-experimenteel onderzoek naar de gevolgen van variaties die in de natuurlijke setting optreden en verder cohortstudies onder representatieve groepen, waarbij retrospectief de prestaties van de kinderen worden vergeleken die wel of geen interventie hebben gehad en waarbij gecontroleerd kan worden voor achtergrondvariabelen die de ontwikkeling beïnvloeden.
Patall *et al*. (2010) vinden het jammer dat er nog zo weinig longitudinaal onderzoek is verricht, zodat de effecten van interventies op lange termijn veelal nog onbekend zijn. Burger (2010) sluit zich hierbij aan. In Nederland hebben Mulder *et al*. (2012) met hun quasi-experiment wel een stap in die richting gezet door de invloed van schakelklassen op de bestrijding van onderwijsachterstanden over een periode van zes jaar te monitoren.

Onderzoekspopulatie
Clark *et al*. (2005) geven te kennen dat het gebruik maken van verschillende respondentengroepen (zoals leerlingen, ouders en leerkrachten) goed werkt.
In de systematische review van Patall *et al*. (2010) laten zij zien dat het kwalijk kan zijn wanneer er op verschillende (abstractie)niveaus met onderzoekspersonen wordt omgegaan. Ze beschrijven een voorbeeld van het schoolsgewijs toewijzen van een bepaalde interventie, maar de uitkomstmaat op individueel leerlingniveau te meten. Ook wijzen ze op de veelal kleine steekproeven, waarop de onderzoeksresultaten worden gebaseerd (Patall *et al*., 2010; Meyer & Van Klaveren, 2011).
Lauer *et al.* (2006) vinden in hun meta-analyse wel goede studies, maar hebben er moeite mee dat *at risk* leerlingen onderling met elkaar vergeleken worden, weliswaar keurig in een experimentele en een controlegroep. Gevonden effecten zeggen dan niets over of de interventie in staat is de kloof tussen achterstandsleerlingen en bevoorrechte leerlingen te verkleinen.
Muijs *et al*. (2004) hebben een literatuurreview verricht en beschrijven een Amerikaans project dat gericht was op *disadvantaged children*. De interventie bleek alleen positieve effecten te hebben bij een subgroep bestaande uit de relatief minst benadeelde leerlingen. Hierdoor is brede inzetbaarheid van zo'n interventie dubieus.
Aan het einde van haar literatuurreview over de effectiviteit van *afterschool programs* concludeert Kassenberg (2010) dat één van de verklaringen waarom sommige studies geen effecten vinden, gelegen is in de heterogeniteit van de onderzoekspopulatie. De enorme variatie tussen schoolkinderen kan leiden tot een type-II-fout, waarbij een in werkelijkheid wel aanwezig verschil tussen experimentele

en controlegroep ten onrechte niet gevonden wordt. Patall *et al*. (2010) waarschuwen juist voor type-I-fouten, waarbij door een veelheid van analysen altijd wel een paar significante effecten gevonden worden, die er in realiteit niet zijn (bij een *p*-waarde van .05 mag alleen al op basis van kansberekening een oneigenlijk effect verwacht worden bij één op de twintig onderzochte relaties; dit wordt wel de *fishing expedition* genoemd).

Verstorende variabelen

Een onderzoeksdesign dat weliswaar keurig gebruik maakt van een controlegroep, maar dat vervolgens ruimte laat voor besmetting van die controlegroep met de interventie, kan als zwak gekarakteriseerd worden. Hiervan is in wezen ook sprake als een programma op werkzaamheid getoetst wordt, maar er allerlei situationele factoren (oncontroleerbaar) meespelen (politieke context, sociaal welzijn of arbeidsmarkt) die van invloed zijn op onderzoekspersonen of hun directe omgeving (Geddes *et al*., 2011). Ook Patall *et al*. (2010) worden in hun systematische review over schooltijdverlenging geconfronteerd met *confounders*: factoren, die naast de gemeten onafhankelijke variabelen, een niet van de interventie te onderscheiden invloed uitoefenen op de uitkomstmaat. Kassenberg (2010) raadt aan in de statistische analysen voor allerlei andere variabelen te controleren, waaronder de frequentie en het doel van *afterschool* activiteiten.

In onderzoek naar de effecten van *out-of-school-time* is nooit helemaal te voorkomen dat de controlegroep onzuiver wordt (Lauer *et al*., 2006; Kassenberg, 2010); immers: een leerling die niet aan een dergelijk programma wordt toegewezen, gaat na de reguliere schooltijd misschien naar muziekles, volgt een cursus of werkt via sport aan z'n zelfvertrouwen, allemaal factoren waarop de onderzoeker geen grip heeft. Wie weet eigent zo'n controleleerling zichzelf een pseudo-interventie toe, waardoor het contrast met de experimentele groep uiteraard vermindert. Ook Biedinger (2011) onderkent dit probleem in haar analyse van de thuissituatie: worden kinderen door hun ouders thuis op een leuke manier uitgedaagd om te leren (leerzame uitjes, *quality time* met de ouders) of niet; dit is nauwelijks in kaart te brengen, omdat het zich achter de voordeur afspeelt. Biedinger (2011) verrichtte een secundaire analyse op data verzameld bij 3–4-jarigen (*kindergarten*). Daaruit komt onder andere naar voren dat men beter voor eerder opgedane ervaringen kan controleren dan voor het minder belangrijke opleidingsniveau van de ouders of voor de thuissituatie. Een longitudinaal design stelt eerder in staat voor '*prior achievements*' te controleren.

In de meta-analyse van Van Ewijk & Sleegers (2010) van onderzoeken die gaan over de mate van witheid van schoolklassen, breken de auteurs een lans voor de *Instrumental Variables approach* of *Fixed Effects analysis*. Hiermee kan de onderzoeker voor eventueel vergeten of onbetrouwbaar te meten *confounding variables* corrigeren. Ze concluderen dat het raadzaam is altijd sociaaleconomische status, '*prior achievements*' en '*ability*' als covariaten in analysen op te voeren.

Interventie

Geddes *et al*. (2011) signaleren in hun literatuurreview dat interventies soms onvoldoende gestandaardiseerd zijn. Daardoor is onduidelijk hoe de interventie precies uitgevoerd wordt in verschillende settingen en is dientengevolge moeilijk eenduidig te analyseren en te concluderen of de interventie als effectief bestempeld dient te worden. Nap-Kolhoff *et al*. (2008) en Kassenberg (2010) hebben het over de kwaliteit van zowel de interventie als de uitvoering, die in effectiviteitsanalysen meegenomen zouden moeten worden.

Clark *et al*. (2005) en Muijs *et al*. (2004) signaleren dat de dosering van de interventie zelden in het onderzoeksverslag wordt vastgelegd en zij zien dat als een manco. Hoewel 'dosering' erg medisch klinkt, kan in het verband van deze literatuurstudie gedacht worden aan het aantal uren per week blootstelling (i.e. meedoen) aan een bepaald programma. De dosering wordt veelal niet gemanipuleerd, waardoor bijvoorbeeld de optimale participatiegraad niet vastgesteld kan worden (Roth *et al*., *2010;* Patall *et al*., 2010). Ook Kassenberg (2010) ziet dat in haar literatuurreview in een door haar beschreven studie terug (ID-37).

Lauer *et al*. (2006) constateren dat, omdat *out-of-school-time* direct volgt op de reguliere schooluren, effecten ervan op de cognitieve leerprestaties van leerlingen minder eenduidig te interpreteren zijn.

De onderzoeken die What Works Clearinghouse meta-analyseert, tonen aan of een bepaald leesprogramma al dan niet effectief is. Er kunnen echter geen uitspraken gedaan worden over onderdelen van zo'n programma, zodat onduidelijk blijft of het werkzame ingrediënt niet zozeer het

aantrekkelijke, kindgerichte materiaal is als wel bijvoorbeeld de intensieve omgang tussen leerling en leerkracht (WWC, 2008, 2012).

Uitkomstmaat
Er wordt in de literatuur geconstateerd dat de afhankelijke variabele niet altijd duidelijk gedefinieerd wordt – of op z'n minst niet eenduidig beschreven in het onderzoeksartikel (Geddes *et al*., 2011). Als hiervan sprake is, is niet of moeilijk vast te stellen of de interventie effect sorteert, omdat waaruit zo'n effect zou moeten blijken, onvoldoende geoperationaliseerd is. Een ander type zwakke effectmaat is één, die wel duidelijk omschreven staat en meetbaar is, maar die niet 'past' bij effectiviteitsonderzoek. Een voorbeeld daarvan is oudertevredenheid. Dat ouders al dan niet tevreden zijn over een bepaalde maatregel geeft niet aan of die maatregel effectief is met het gestelde doel voor ogen. Het kan wellicht eerder de acceptatiegraad verklaren dan doeltreffendheid. Zo onderzochten Burgess & Dixon (2012) welke rol lage sociale status heeft op de *perceptie* van privéscholen versus staatsscholen in Columbia. Het analyseren van opvattingen van respondenten staat niet gelijk aan het vaststellen van effectiviteit.
Ook komt het voor dat er te vroeg geëvalueerd wordt (Roth *et al*., 2010; Geddes *et al*., 2011; Meyer & Van Klaveren, 2011). Er wordt dan een effectmeting gedaan op een moment waarop het effect van de interventie nog onvoldoende heeft kunnen 'rijpen'. Dit kan een verklaring zijn voor het vaak niet vinden van een significant effect, zoals gezien in de onderzoeksliteratuur, die op zichzelf waarschijnlijk een onderschatting is van het aantal studies dat geen effecten vond (publicatie *bias*). Longitudinale studies worden door Burger (2010) in diens review dan ook gerekend tot de betere onderzoeksdesigns.
Als het '*monitoring system*' tekort schiet, wordt het vinden van effecten of het doen van betrouwbare uitspraken daarover, bemoeilijkt (Burger, 2010; Geddes *et al*., 2011). Kassenberg (2010) beschrijft in het kort een aantal (overigens niet in een bibliografie opgenomen) studies die gaan over de invloed van *afterschool programs* op de cognitieve talentontwikkeling van leerlingen op het basisonderwijs tot en met het mbo. In één van de beschreven studies (ID-124) wordt geadviseerd gebruik te maken van gestandaardiseerde toetsinstrumenten om de uitkomstmaat verantwoord te kunnen bepalen.

Setting
Diverse auteurs erkennen dat gevonden effecten in bijvoorbeeld Amerikaans onderzoek niet zonder meer geëxtrapoleerd kunnen worden naar bijvoorbeeld de Nederlandse onderwijssituatie of naar andere OECD-landen (Geddes *et al*., 2011; Faubert, 2012; Van Schooten, 2012). Ook de Duitse onderzoeker Biedinger (2011) loopt hier tegenaan. Het speelt ook bij een onderzoek naar Columbiaanse privé- en staatscholen (Burgess & Dixon, 2012), waarvan de resultaten – los van methodologische redenen – weinig implicaties hebben voor de Nederlandse situatie.
Muijs *et al*. (2004) constateren dat zwakpresterende scholen niet of minder open staan voor onderzoek in het algemeen en voor openbaarmaking van onderzoeksresultaten in het bijzonder. Scholen zouden dus meer begeleid moeten worden in de acceptatie van onderzoeksresultaten. Hier is een rol weggelegd voor onderzoekers om doelbewust het onderzoeks*proces* meer aandacht te geven.

5.3 Overwegingen voor keuze onderzoeksdesign

Diverse auteurs, van wie artikelen gelezen werden in het kader van hoofdstuk 2 tot en met 4, hebben in hun discussieparagraaf en soms in de beschrijving van hun methoden aandachtspunten beschreven met betrekking tot de onderzoeksmethodologie. In de vorige paragraaf is daarvan verslag gedaan. In deze paragraaf wordt geprobeerd lering daaruit te trekken.

Zonder twijfel is een zuiver experiment de beste methode om de werkzaamheid van een bepaalde interventie (statistisch) aan te tonen. Verder voldoet het 'ideale' onderzoeksdesign aan de volgende voorwaarden:

- de kwantitatieve onderzoeksresultaten worden verrijkt met behulp van andere onderzoeksmethoden, waaronder kwalitatieve;
- de afhankelijke variabele wordt eenduidig gedocumenteerd en goed geoperationaliseerd; hierbij wordt gebruik gemaakt van gestandaardiseerde en gevalideerde meet- of toetsinstrumenten;
- de data wordt op verschillende manieren verzameld;

- de steekproeven zijn van voldoende omvang (representativiteit, *power*berekeningen) om verantwoorde analysen te kunnen verrichten en gerechtvaardigde conclusies te kunnen trekken;
- verschillende gebruikersgroepen van de interventie worden in het onderzoek betrokken, die per subgroep redelijk homogeen zijn;
- de selectie van onderzoekssubjecten past bij het niveau van analyse;
- de interventie is gestandaardiseerd (geprotocolleerd), waarbij er zo weinig mogelijk (trainergebonden) variatie bestaat in de uitvoering van de interventie of het programma;
- de interventie wordt in verschillende maten van 'sterkte' aangeboden aan de experimentele subgroepen (dosering, veelal op te vatten als frequentie en deelnametrouw);
- ook op de merites van specifieke interventieonderdelen kan geanalyseerd worden;
- op grond van meetbare, te verbijzonderen onafhankelijke variabelen kunnen subgroepanalysen verricht worden;
- de condities voor de controlegroep kunnen zo goed mogelijk beheerst worden, zodat er geen contaminatie-effecten kunnen ontstaan, waarvoor in de analysefase niet gecontroleerd zou kunnen worden;
- 'alle' *confounders* worden geïdentificeerd en gemeten en voor hun invloed op het interventie-effect wordt gecorrigeerd; sowieso moeten variabelen als *prior achievements*, sociaaleconomische status en *ability* als covariaten meegenomen worden;
- er is tijd de experimentele en controlegroep gedurende langere tijd te volgen, zodat er voldoende tijd is voor het optreden van het effect, maar ook eventuele uitdoving van een effect zichtbaar kan worden; longitudinaal onderzoek met herhaalde metingen is daarvoor geëigend;
- er moet voldoende aandacht zijn voor het onderzoeksproces (werving, trouw, voorkoming van uitval, aandacht voor sleutelfiguren en respondenten).

Spanningsveld

Het bovenbeschreven design mag dan wel een ideaaltype zijn, er moet ook rekening gehouden worden met een aantal randvoorwaarden, zoals tijd, geld en motivatie. Langdurig toegepast onderzoek vraagt veel tijd die er vaak niet is. Bovendien is het kostbaar. Ten slotte vraagt het van scholen de bereidheid om lange tijd aan een onderzoek deel te nemen en gegevens aan te leveren. Uitbreiding van de onderzoekspopulatie (diverse gebruikersgroepen, beredeneerde subgroepen daarbinnen en al die groepen in een verantwoord grote steekproef en gematchte controlegroep) doet de kosten logaritmisch stijgen. En dit alles – in de onderwijssetting – nog los van de veelal praktische onmogelijkheid op leerlingniveau te randomiseren.

Er bestaan zowel goedkopere als snellere vormen van onderzoek. Bij secundaire analysen is de data al verzameld en kan de onderzoeker meteen aan de slag. Groot nadeel is natuurlijk, dat die data met het oog op andere onderzoeksvragen verzameld is en de onderzoeker geen vat heeft op de operationalisering van de variabelen of nog voor niet-gemeten *confounders* kan corrigeren. Retrospectief onderzoek, waarbij de interventie in het verleden gepleegd is en de effecten op een later tijdstip gemeten worden, is weliswaar tijdsefficiënt (niet wachten op een effect), maar vormt in sommige gevallen een onontwarbare kluwen van beïnvloedende factoren (covariaten), met hun onbekende (want niet gemeten) invloed op de effectmaat. Verder is er achteraf geen voldoende zicht meer te verkrijgen op hoe de interventie is toegediend en is er geen controlegroep. Dit alles moet met het oog op bewijskrachtige onderzoeksdesigns als zwakten betiteld worden. Maar kostenbewust(er) en snel(ler) zijn deze vormen wel, net als dossieronderzoek, waarvoor in grote lijnen dezelfde voor- en nadelen gelden.

Het verrichten van uitsluitend kwalitatief onderzoek, bijvoorbeeld in de vorm van interviews, individueel of in groepen, of (participerende) observaties is vanwege de bewerkelijkheid ervan veelal duur, maar kent ook het probleem van de representativiteit en extrapoleerbaarheid.

Alternatieven

Dit spanningsveld vraagt om een optimaal (i.e. maximaal onder de gegeven omstandigheden) onderzoeksdesign. Het Sociaal en Cultureel Planbureau heeft zeer recentelijk allerlei beleidsmaatregelen voor het onderwijs tegen het licht gehouden (Herweijer & Turkenburg, 2012). Dit wetenschappelijk bureau constateert ook dat ideaaltypisch onderzoek in de praktijk vanwege (methodologische, beleids-inhoudelijke, ethische of randvoorwaardelijke) redenen veelal niet haalbaar is. Zij komen met een

aantal secondbest oplossingen, die hoofdzakelijk erop neer komen handig om te gaan met het formeren van een controlegroep.

Zo kan het probleem van de randomisering op leerlingniveau 'opgelost' worden door dat op groepsniveau te doen. Niet de leerling wordt op basis van toeval toegewezen aan de experimentele dan wel controlegroep, maar een hele klas of een hele school. Het is voorstelbaar dat het in de praktijk onmogelijk is de helft van de leerlingen van een klas of een paar klassen van een school de interventie toe te dienen en een ander gedeelte niet. Nog ongeacht de organisatorische valkuilen voor de school of de leerkracht en wellicht ethische overwegingen (het een groep onthouden van een mogelijk profijtelijke interventie), is het gevaar van contaminatie tussen beide onderzoeksgroepen levensgroot. Groepsgewijze toewijzing kan dan een uitkomst zijn, hoewel in § 5.2 bleek dat Patall *et al*. (2010) hierbij kanttekeningen plaatsen. Onderzoeksmethodologen waarschuwen voor de zogenaamde 'ecologische valkuil', waarbij op groepsniveau gevonden resultaten niet zonder meer overzetbaar zijn naar individuen van die groep.

Verder kan een beleidsmaatregel weliswaar integraal (landelijk) worden ingevoerd, maar als dat in fasen gebeurt, dan kan een subgroep die later aan de beurt is tijdelijk als controlegroep voor de eerste experimentele groep dienen. Het SCP waarschuwt ervoor dat de fasering van de implementatie dan niet mag samenhangen met de ernst van de problematiek of een andere aan de (on)afhankelijke variabelen gecorreleerde factor. Een controlegroep kan ook geconstrueerd worden wanneer de beleidsmaatregel de ene groep wel en de andere veel minder treft; de verschillende mate van beleidsintensiteit maakt dan een onderscheid mogelijk (Herweijer & Turkenburg, 2012).

Een andere elegante oplossing voor het formeren van een controlegroep is in de effectmeting niet te kijken naar absolute scores, maar naar behaalde winst. Ook Van Ewijk & Sleegers (2010) hebben het over *gain scores* versus *present scores*. Leesvaardigheid zal bij bijvoorbeeld autochtone kinderen op hogere scores uitkomen dan bij allochtone leerlingen, maar er kan ook gekeken worden naar de winstpercentages (hoeveel (procent)punten progressie geboekt in leerjaar *x*?). Onder de aanname dat de stijgingstrend in de controlegroep vergelijkbaar zou zijn geweest aan die in de experimentele groep, ware het niet dat daarin een interventie is gepleegd, wordt de extra groei aangemerkt als een bewijs in de richting van een effectieve maatregel. Dit wordt in de wetenschappelijke methodologie de *difference-in-difference*-benadering genoemd.

Net als hierboven al aangehaald werd, wijst het SCP (Herweijer & Turkenburg, 2012) op de mogelijkheden van matching (voor elke proefpersoon in de experimentele groep een evenknie zoeken voor in de controlegroep, zodat op variabelen uit het theoretische model beide onderzoeksgroepen vergelijkbaar zijn) en statistische controle voor *confounders*. Ook zij vermelden de tekortkomingen daarvan: het is nooit bekend of op alle wezenlijke variabelen wel gematcht is dan wel of daarvoor in regressiemodellen gecontroleerd is.

Hoe meer premissen moeten worden gemaakt en hoe minder zicht er is op de geldigheid van die aannamen, des te zwakker worden de onderzoeksdesigns, omdat het steeds moeilijker wordt om gevonden effecten eenduidig toe te schrijven aan de interventie. Sowieso moeten onderzoekers bescheiden zijn als ze conclusies willen trekken (zie ook Patall *et al*., 2010; § 5.2): bijvoorbeeld in het geval van toepassing van de *difference-in-difference*-benadering in een design waarbij de controlegroep een geringe mate van beleidsintensiteit ervaart en de experimentele groep juist een grote mate, kan de conclusie niet zijn dat de beleidsmaatregel effectief is, maar 'slechts' dat er een verschil is in de effectiviteit, al naar gelang de beleidsintensiteit.

Het SCP noemt tenslotte nog een andere optie, die handig omgaat in geval van het ontbreken van een controlegroep. Stel dat een trend ten aanzien van bepaalde scores of gedrag van een zekere populatie vanuit het verleden bekend is, dan kan die trendlijn doorgetrokken worden naar het heden. Als vervolgens metingen worden gedaan en de resultaten daarvan blijken af te wijken van de verwachte trend, dat zou dat verschil aan de interventie toegeschreven kunnen worden. Het zal duidelijk zijn dat hieraan nogal wat mitsen en maren kleven. Net als bij retrospectief onderzoek is niet duidelijk welke verstorende variabelen al dan niet gelijktijdig met de interventie zich hebben voorgedaan. En uiteraard staat of valt deze optie met de validiteit van het voorspellingsmodel rond de trend (Herweijer & Turkenburg, 2012).

Onderzoeksdesigns die effectieve bestrijding onderwijsachterstanden aantonen
Omdat ervoor gekozen is in principe alleen systematische reviews en meta-analysen als bronnen te raadplegen (zie § 1.3), is er niet zoveel zicht op hoe de onderliggende onderzoeken uitgevoerd zijn.
Een tweede bron bestaat uit recente artikelen (2011 of 2012), die waarschijnlijk nog niet in meta-studies of reviews opgenomen kunnen zijn. Vanwege de zeggingskracht van dergelijke eigenstandige onderzoeken gaat de aandacht vooral uit naar (quasi) experimentele designs. Uiteraard vormt deze handelwijze al een richtinggevende voorselectie. Een derde stroom van literatuur is verworven via de sneeuwbalmethode.

Om een indruk te geven hoe in onderzoeksdesigns elementen van robuuste designs terugkomen, volgen hierna twee voorbeelden; het eerste betreft het correlationele onderzoek van Burgess & Dixon (2012), het tweede het quasi-experiment van Mulder *et al*. (2012).
Burgess & Dixon (2012) onderzochten het verschil in effectiviteit tussen particuliere en staatsscholen in achterstandsgebieden van Columbia. Ze maakten gebruik van twee gebruikersgroepen (leraren en ouders), die ze elk een gevalideerde vragenlijst voorlegden, waarna er ongestructureerde interviews met respondenten plaatsvonden (mix van onderzoeksmethoden). Helaas konden de items van elk van de vragenlijsten niet met elkaar vergeleken worden. De gewenste matching op basis van schoolgrootte mislukte, omdat alle staatsscholen minstens drie keer zo groot waren dan particuliere scholen. Om het aantal covariaten overzichtelijk te houden, kozen de onderzoekers voor een subselectie van een bepaald schooltype, waardoor meer dan de helft van de scholen afviel. Daarbinnen werd op willekeurige wijze de helft van die scholen in het uiteindelijke onderzoek betrokken. De metingen werden verder op individueel niveau verricht.

Het enige gevonden quasi-experiment dat over onderwijsachterstandenbestrijding in het primair onderwijs gaat is dat van Mulder *et al*. (2012), nota bene een Nederlands onderzoek. Het gaat over het effect van schakelklassen op cognitieve vaardigheden (taal-, lees- en rekenprestaties) en sociale competenties (zelfvertrouwen, gedrag en schoolwelbevinden, maar die komen in het onderzoeksverslag vreemd genoeg verder niet meer terug). Schakelklassen zijn er in soorten en maten: naschoolse uren een aantal jaren lang, een heel leerjaar aan het begin van de basisschool (eigenlijk voorafgaand), een heel schooljaar ergens halverwege de basisschool en een heel schooljaar na groep 8, ter voorbereiding op het voortgezet onderwijs (kopklas). Kopklassers waren deels hun eigen controlegroep (hun prestaties in groep 8 werden vergeleken met die van na het daaropvolgende schakeljaar). Voor de rest bestond de controlegroep uit leerlingen die (uiteraard) geen schakelklas volgden, maar wel een vergelijkbare score op de Cito-eindtoets haalden, vergelijkbare leercapaciteiten bezaten en een vergelijkbare etnische achtergrond hadden.
Leerlingen werden vier jaar lang gevolgd: na de nulmeting volgden nog twee nametingen, waarbij de onderzoekers in voorkomende gevallen contact hielden met scholieren tot op het voortgezet onderwijs. De veelal allochtone schakelklasleerlingen boekten méér leerwinst op taal dan vergelijkbare kinderen in de controlegroep. Kopklasleerlingen haalden hogere scores voor hun Cito-eindtoets dan ze zelf een jaar eerder deden en kregen vaker adviezen voor hogere vormen van voortgezet onderwijs dan eerder. Schakelklasleerlingen voelden zich niet gestigmatiseerd omdat ze in een schakelklas zaten. Uiteindelijk behielden schakelklasleerlingen een achterstand op hun klasgenoten, maar dit kon eerder toegeschreven worden aan het grote verschil op instroomniveau; met andere woorden: schakelklasleerlingen liepen een deel van hun achterstand in. Ook op langere termijn bleven deze resultaten nog duidelijk zichtbaar, zij het dat in het geval van onderbouwschakelklassen er na één of twee jaar nadien weinig over leek te zijn van de extra leerwinst op woordenschat en rekenen, maar dat bleek een onderzoeksronde later een tijdelijke dip te zijn. Op begrijpend lezen en technisch lezen was die dip er überhaupt niet. Wat betreft de hoogte van de adviezen voor vervolgonderwijs was er nog maar nauwelijks verschil met controlegroepleerlingen.

Elementen die in § 5.2 en het eerdere van deze paragraaf als robuust naar voren kwamen en die in het onderzoek van Mulder *et al*. (2012) teruggevonden zijn, zijn de volgende:
- er is sprake van een experimentele (schakelklas) en een controlegroep (COOL- en PRIMA-data van reguliere leerlingen); deelname aan de experimentgroep is duidelijk afgebakend;
- er is sprake van *matching* op cruciale, onafhankelijke variabelen;

- de interventie is concreet (extra schooluren);
- de effectmaat, gemeten met gestandaardiseerde instrumenten als Cito-toetsen, is gekoppeld aan leerwinst en niet aan de absolute hoogte van de scores;
- ook andere effectmaten worden gebruikt (hoogte van advies voor voortgezet onderwijs – hoe subjectief ook, toch is dit wel een eenduidige maat –, zitten blijven, afstromen op voortgezet onderwijs, rapportcijfers) en verder percepties (al dan niet gevoel van stigmatisering);
- dataverzamelingsmethoden zijn prospectief dossieronderzoek (scores opvragen) en vragenlijsten;
- er is sprake van longitudinaal onderzoek (vier jaar lang gevolgd); hierdoor werd duidelijk dat de dip in scores niet het weg ebben van het effect van de schakelklas is, maar slechts tijdelijk bleek te zijn;
- de respons was in de deelonderzoeken voor het basisonderwijs rond de 95% en voor het voortgezet onderwijs circa 80%; wel nam de respons af naarmate leerlingen langer gevolgd werden;
- de *n* lijkt voldoende (totaal ± 2.000 schakelklasleerlingen, uitgesplitst over verschillende subgroepen);
- de onderzoeksgroepen zijn redelijk homogeen;
- de selectie als ook de analyse zijn op het niveau van individuele schakelklasleerlingen;
- hoewel er sprake is van andere groepen respondenten (leraren, ouders, gemeenteambtenaren, interne begeleiders, schoolleiders) wordt hierover niet gerapporteerd, anders dan dat men enthousiast en gemotiveerd is.

Wat in Mulders onderzoek ronduit zwak is, is dat de aard van de interventie niet nader gepreciseerd wordt: er wordt slechts gesteld dat schakelklassen intensief taalonderwijs betreffen. Er is dan ook geen onderzoek gedaan naar verschillende doseringen van de interventie (duur van schakelklas). Hoewel schakelklassen ook in de vorm van een extra tussenjaar gestalte kunnen krijgen, lijkt het onderzoek voornamelijk op parallelklassen gericht, op de kopklassen na, die een extra schooljaar na groep 8 betekenen. Verder is er sprake van een tweejarige schakelklas, maar wordt deze subgroep vanwege een te geringe omvang niet meegenomen in het onderzoek. Ook wordt niet gekeken naar de uitkomsten van de verschillende varianten: leveren bijvoorbeeld onderbouwschakelklassen betere resultaten (op langere termijn) op dan schakelklassen in de middenbouw? En omdat de interventie inhoudelijk niet toegelicht wordt, is er niets te zeggen over de effectiviteit van bepaalde onderdelen.
Een andere relatieve zwakte betreft de schakelklasvariant waarin gedurende een aantal jaren een paar uur per week naschoolse uren worden aangeboden. Het is de vraag in hoeverre er andere factoren meespelen in de behaalde resultaten. Omdat het over verscheidene jaren uitgesmeerd is, is er een grotere periode waarin allerlei verstorende variabelen kunnen optreden, nog los van dat het moeilijk is de extra leerwinst te onderscheiden van het reguliere onderwijs. In hoeverre de interventie geprotocolleerd is, wordt uit het onderzoeksverslag niet helemaal duidelijk.
De onderzoekers constateren een sterke oververtegenwoordiging van onderbouwschakelklassen ten opzichte van middenbouw-, bovenbouw- of kopschakelklassen, maar doen hiermee niets in hun analysen, omdat ze de leerlingen volgen en alleen per subgroep uitspraken doen. Toch tast deze handelwijze de generaliseerbaarheid van de onderzoeksresultaten aan.
Ook wordt neveninput tijdens de schakelklas (bijvoorbeeld meer individuele aandacht voor de taalzwakke leerling) niet nader onderzocht. In hun discussiehoofdstuk geven de onderzoekers zelf verder aan dat onbekend is of schakelklasleerlingen in de periode daarna nog extra begeleiding ontvingen.
De onderzoekers werden geconfronteerd met een tussentijdse wijziging van de Cito-systematiek. Daardoor kon er geen onderscheid meer gemaakt worden tussen woordenschat, begrijpend lezen en technisch lezen. Dit is een kwestie van overmacht. Het was prettiger geweest als men zelf de regie over de onderzoeksinstrumenten had kunnen hebben. Wellicht hadden de onderzoekers de oude versie van de Cito-instrumenten nog kunnen inzetten, hoewel dat wel een extra belasting zou impliceren.
Omdat de controlegroep uit *gematchte* leerlingen bestaat voor wie de resultaten via secundaire analyse verkregen worden, is het de (niet te beantwoorden) vraag of wel voor alle van belang zijnde aspecten gecontroleerd is. Het onderzoeksverslag maakt geen melding van *confounders*. Hoewel het onderzoek longitudinaal van aard is, lijkt er niet voor *prior achievements* gecorrigeerd te worden.

Mulder *et al*. (2012) verrijken hun kwantitatieve onderzoeksresultaten nauwelijks (in elk geval niet gedocumenteerd) met andere vormen van onderzoek; uiteraard heeft dit alles met de aard van de onderzoeksvragen te maken.

Synthese
Dit alles overziend komen de onderzoekers tot de volgende voorstellen, die in § 5.4 worden gepresenteerd.

5.4 Voorgestelde onderzoeksdesigns

ONDERZOEKSDESIGN A
Het Ministerie van Onderwijs, Cultuur en Wetenschappen heeft zelf een voorzet gedaan voor een design, waarin in het volgende voorzien wordt. Van scholen met een vergelijkbare achterstandspopulatie worden de Cito-scores in kaart gebracht. Tevens worden de gepleegde interventies van deze scholen in kaart gebracht. Op basis van deze gegevens worden vervolgens de faal- en succesfactoren bepaald. Met andere woorden: welke interventies correleren met een hoge dan wel lage score op de Cito-toets. Vervolgens wordt laagscorende scholen gevraagd om te stoppen met de interventies die correleren met lage scores op de Cito-toets en de vrijgekomen middelen in te zetten voor interventies die correleren met hoge scores op de Cito-toets. Een toename in de Cito-score zou dan een aanwijzing zijn dat de gepleegde interventies effectief zijn.
Deze onderzoeksopzet heeft veel weg van wat Muijs *et al*. (2004) in hun *narrative* review hebben gedaan (zie hoofdstuk 3).

ONDERZOEKSDESIGN B
1. Via een paar focusgroepsbijeenkomsten / expertmeetings van deskundigen (beleidsmakers van het Ministerie van Onderwijs, stafmedewerkers van Onderwijsbonden, leerplichtambtenaren) en betrokkenen (leraren, interne begeleiders) wordt achterhaald hoe door hen de acceptatiegraad van verschillende interventies gepercipieerd wordt. Deze interventies zijn de in hoofdstuk 2 tot en met 4 geïnventariseerde elders effectief gebleken programma's ter bestrijding van onderwijsachterstanden, voor zover scholen die zelf kunnen implementeren.
2. De drie interventies worden in extenso beschreven: onder welke condities worden ze uitgevoerd, wat zijn kwaliteitscriteria en hoe kunnen die geborgd en gecheckt worden?
3. Het CPB wordt gevraagd een indicatie te geven van de prijskaartjes die aan de vijf door het Nederlandse onderwijsveld meest geaccepteerde interventies hangen mede op grond van de uitgebreide interventiebeschrijvingen. De drie goedkoopste interventies gaan door voor verder onderzoek.
4. De uitkomstmaat kan zowel cognitief als sociaal-emotioneel zijn. Nadat deze geëxpliciteerd zijn – ook hiervoor worden de experts ingezet –, worden hiervoor gevalideerde meetinstrumenten gezocht of ontwikkeld. Het kan gaan om toetsresultaten (*hard skills*), percepties of gedragingen (*soft skills*). Raadpleging van administratiesystemen (Magister, SOM) is handig om toetsresultaten te achterhalen; vragenlijstonderzoek is behulpzaam bij het bepalen van overige effectmaten.
5. Omdat prospectief onderzoek meer handvatten geeft *confounders* te elimineren of anders bijtijds voor de analysefase te meten, en longitudinaal onderzoek het effectverloop op langere termijn in kaart brengt, wordt een dergelijk quasi-experiment voorgesteld.
6. Aanvullende literatuurstudie of expertmeetings (wellicht met de ontwikkelaars van een bepaalde programma) aangaande de drie interventies levert mogelijke covariaten op. Hierbij moeten meetinstrumenten gezocht worden, die deze variabelen op valide wijze meten.
7. Er worden een *x*-tal experimentgroepen geformeerd, waarbij *x* staat voor het aantal doseringen (uurcategorieën) waarin de interventie kan worden toegediend, vermenigvuldigd met 3 (door het onderwijsveld geaccepteerde interventies), vermenigvuldigd met 3 (het aantal te onderzoeken onderwijstypen: basisonderwijs, speciaal basisonderwijs en (voortgezet) speciaal onderwijs). Bij een startaantal van 100 leerlingen per subgroep zou dan een experimentgroep in totaal 2.700 leerlingen omvatten, uitgaande van drie doseringsgradaties of deelvarianten.

8. De controlegroep bestaat uit *gematchte* leerlingen. Omdat dezen juist niet een bepaalde interventie ontvangen, kan volstaan worden met 100 leerlingen per onderwijssetting, dus in totaal 300 controleleerlingen.
9. Overwogen kan worden gebruik te maken van twéé controlegroepen: één om de resultaten van interventieleerlingen af te zetten tegen relevante kenmerken vergelijkbare leerlingen en een tweede om hetzelfde te doen maar dan bij de algemene groep klasgenoten. De eerste controlegroep stelt in staat de leerwinst te bepalen vanwege de gepleegde interventie (relatieve scores), terwijl de tweede nodig is om te beoordelen of onderwijsachterstanden daadwerkelijk worden ingelopen (absolute scores).
10. Ook andere betrokkenen – naast leerlingen – worden in het onderzoek opgenomen. Te denken valt aan meningen van leerkrachten, interne begeleiders en schoolleiders omtrent de uitvoerbaarheid en geschiktheid van de interventie. Deze informatie wordt in een later stadium afgezet tegen de primaire effectmaten.
11. Onderzoeksresultaten worden door eerder geraadpleegde experts in een inhoudelijk kader geplaatst, geduid en van kanttekeningen voorzien. Voor achtergrondinformatie betreffende specifieke resultaten zijn diepte-interviews met experts of gebruikers geëigend.
12. Er wordt veel tijd en aandacht gestoken in het onderhouden van relaties met deelnemers aan de onderzoeken, zodat afspraken (bijvoorbeeld over de concrete uitvoering van de interventie) zo goed mogelijk nagekomen worden en de respons optimaal is.

ONDERZOEKSDESIGN C

Leseman (2007) stelt dat in een evaluatiedesign de volgende onderdelen niet mogen ontbreken: (a) de kwaliteit van de vve, (b) de inhoud van de interventies en (c) de continuïteit van de interventie, vertaald in de integratie met een programma voor ouders, samenwerking met andere instellingen en een goede aansluiting met het primair onderwijs.

In de VS zijn goede programma's zeer intensief en langdurig. De programma's zijn nauwelijks vergelijkbaar met het Nederlands aanbod. Als scholen goede resultaten bereiken met het vve-programma, is het belangrijk om te onderzoeken hoe zij het vve-programma gestalte geven. Deze '*lessons-to-learn*' kunnen worden meegenomen in een vervolgonderzoek.

Dit onderzoeksdesign voorziet, in navolging van het EPPE-project, in de bestudering van twaalf cases. Op basis van bestaande data (COOL, DUO, Cito) wordt gezocht naar twee groepen scholen met een groot aantal gewichtenleerlingen. Alle scholen hebben een aanbod op het gebied van onderwijsachterstanden:

- zes scholen met gemiddelde tot goede Cito-scores op taal en rekenen van leerlingen in diverse leerjaren;
- zes scholen met benedengemiddelde Cito-scores op taal en rekenen van leerlingen in diverse leerjaren.

De scholen worden bestudeerd door middel van observatie, checklists en interviews aangaande de vraag wat de werkzame ingrediënten van de succesvolle scholen zijn. Waarin verschillen de scholen en de uitvoering van de onderwijsachterstandsinterventies van elkaar? Er wordt gelet op kwaliteit, intensiteit, instroom leerlingen, *monitoring* van de leerlingen en gebruik van meetinstrumenten en verder de kenmerken bekend uit de (internationale) literatuur, zoals in de schema's van hoofdstuk 2, 3 en 4 opgenomen (zie voor een synthese daarvan in § 5.6).

Door middel van kwalitatief onderzoek kunnen 'witte vlekken' met betrekking tot het succesfactoren aan het licht komen.

Deze succesfactoren kunnen worden ingezet in een vervolgonderzoek onder een groot aantal basisscholen waarbij op prospectieve wijze gegevens worden verzameld.

5.5 Intermezzo: gebruik van reeds verzamelde data

5.5.1 Inleiding

In een voorbespreking met de opdrachtgever is de voorkeur voor onderzoeksdesign C (zie § 5.4) uitgesproken. In deze onderzoeksopzet wordt voorgesteld gebruik te maken van landelijk verzamelde

gegevens van PRIMA, COOL en DUO. SEOR is gevraagd om te kijken hoe bruikbaar deze gegevens zouden kunnen zijn. In deze paragraaf volgt een verslag van deze expeditie.
Allereerst worden de mogelijkheden van bestaande databestanden besproken (a) voor het meten van effecten van onderwijsachterstandenbeleid op de leerprestaties van leerlingen in het primair onderwijs en (b) voor het onderzoeksdesign van toekomstig effectiviteitsonderzoek.
Achtereenvolgens wordt beschreven de COOL-data (inclusief PRIMA, de voorloper ervan), data beschikbaar via DUO en data uit de Vve-monitor. Daarbij komen tevens de mogelijkheden voor koppeling van bestanden en de mogelijke meerwaarde daarvan aan de orde. Daarna wordt ingegaan op de hiervoor genoemde vragen: de mogelijkheden voor effectmeting en de rol in het onderzoeksdesign.

5.5.2 PRIMA/COOL

Inhoud van COOL

In Nederland wordt al geruime tijd onderzoek gedaan naar onderwijsloopbanen van kinderen. Hiertoe worden door een breed consortium (waaronder ITS en Kohnstamm) door middel van enquêtes gegevens verzameld van een groot aantal scholen en leerlingen. COOL is de opvolger van de vroegere PRIMA- en VOCL-cohorten. In COOL worden leerlingen in de leeftijd van 5 tot 18 jaar gevolgd in hun schoolloopbaan in het primair onderwijs, voortgezet onderwijs en het Mbo.
In het onderzoek staan de volgende aspecten van de ontwikkeling van leerlingen centraal:

- de cognitieve ontwikkeling: kennis en vaardigheden in het Nederlands, Engels en rekenen/wiskunde;
- de ontwikkeling van sociale competenties, waaronder burgerschapscompetenties;
- de sociaal-emotionele ontwikkeling.

Om de leerlingen op deze drie aspecten te volgen, worden op diverse momenten toetsen en vragenlijsten afgenomen. Daarnaast wordt de gehele schoolloopbaan van de leerlingen in kaart gebracht.
Er zijn drie ronden van gegevensverzameling gepland. In de eerste ronde (2007/2008) hebben leerlingen meegedaan uit groep 2, 5 en 8 van het primair onderwijs en klas 3 van het voortgezet onderwijs. In de tweede en derde ronde (2010/2011 en 2013/2014) worden deze groepen opnieuw onderzocht, maar daarnaast ook de leerlingen in de bovenbouw van het Havo en Vwo en in het Mbo. In elke ronde gaat het om ongeveer 50.000 leerlingen in het primair onderwijs en ongeveer 20.000 leerlingen in leerjaar 3 van het voortgezet onderwijs. In de tweede en derde ronde doen nog eens 20.000 leerlingen verspreid over de examenjaren van Havo en Vwo en leerjaar 2 in het Mbo mee.
De inhoud van PRIMA en COOL is vergelijkbaar in de zin dat dezelfde kenmerken worden uitgevraagd over leerlingen, ouders en scholen. Om een indruk van de inhoud te geven, wordt nader ingegaan op COOL 2010/2011.
COOL 2010/2011 bevat gegevens van in totaal 553 scholen. Hiervan behoren 406 tot de referentiesteekproef en de resterende 147 tot de aanvullende steekproef van scholen met een oververtegenwoordiging van achterstandsleerlingen. De referentiesteekproef beoogt een representatief beeld te geven van de totale populatie scholen in Nederland. Van deze scholen deden er 342 mee aan de vorige meting van COOL (2007/2008). Een deel van de scholen die niet meer wilden meedoen aan COOL 2010/2011, heeft op verzoek wel gegevens aangeleverd over de toetsresultaten van leerlingen die in de vorige meting in groep 2 of 5 zaten.
De deelnemende scholen hebben gegevens aangeleverd over leerlingen in groep 2, 5 en 8. In totaal zijn van circa 38.000 leerlingen gegevens beschikbaar. Het aandeel uit COOL 2007/2008 bekende leerlingen ligt net onder vijftig procent. De uitval wordt veroorzaakt door uitval van scholen, maar ook doordat leerlingen blijven zitten of naar een andere school zijn gegaan. Leerlingen die aan $COOL_{5\text{-}18}$ deelnemen, blijven in het onderzoek betrokken, óók als ze doubleren, van school veranderen of de overstap naar het voortgezet onderwijs maken. Het oorspronkelijke cohort blijft dus (zoveel mogelijk) intact.
COOL bestaat uit verschillende onderdelen die betrekking hebben op leerling, school of ouders. Hieronder wordt in hoofdlijnen ingegaan op de inhoud van COOL.
Leerling gegevens

- geslacht;
- geboortedatum;
- leerling gewicht;
- verblijfsduur in Nederland;

- herkomst ouders;
- opleidingsniveau ouders;
- toets gegevens (diverse toetsen, waaronder Cito);
- subjectieve gegevens over leerling profiel (gedrag, populariteit, werkhouding etc.);
- score op burgerschapskenmerken (attitude, vaardigheid, reflectie en kennis);
- subjectieve scores op welbevinden e.d.

Ouder gegevens

- gezinssamenstelling;
- herkomst ouders/grootouders;
- verblijfsduur in Nederland;
- opleidingsgegevens ouders;
- geloof;
- beheersing (actief en passief) Nederlandse taal;
- interactie kinderen (bijvoorbeeld bezoek bibliotheek, samen tv kijken etc.).

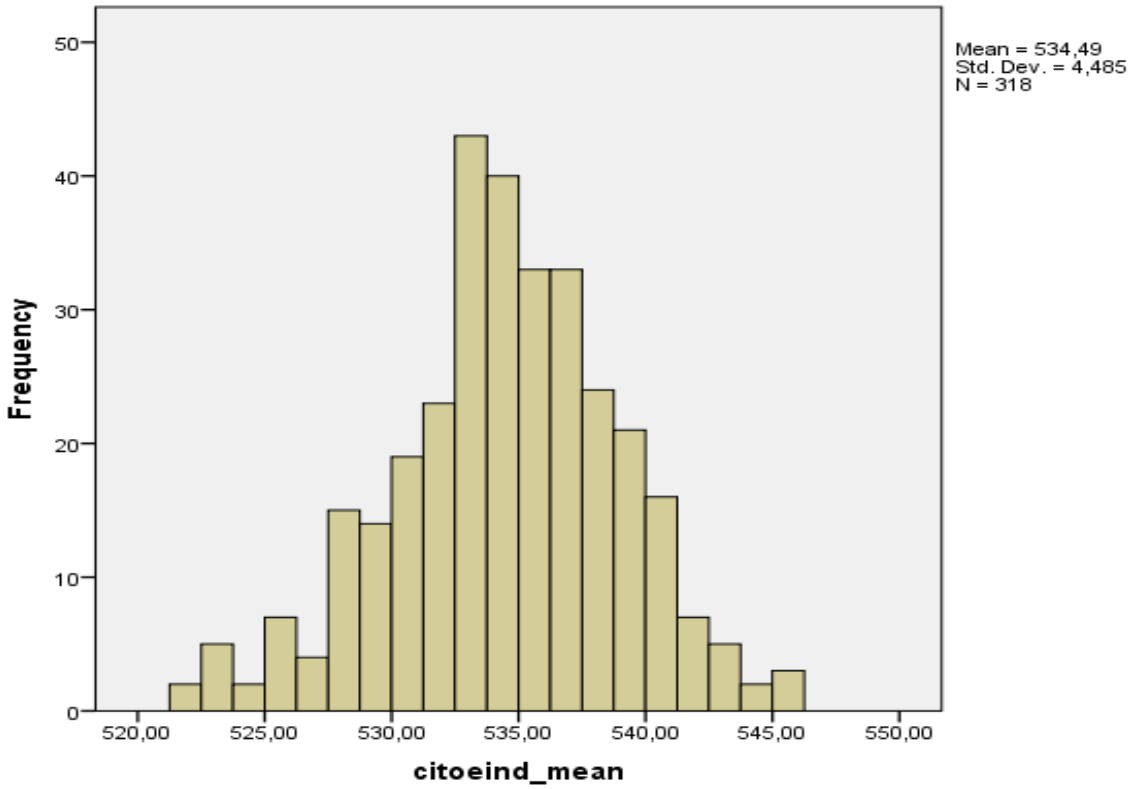
Schoolgegevens

- vestigingsplaats;
- denominatie;
- programma's (vve, integratie peuterspeelzaal etc.);
- schoolscore gewichtenregeling.

Niet alle onderdelen van de hierboven genoemde gegevens wordt uitgevraagd over alle leerlingen. Zo worden gegevens over deelname van leerlingen aan vve alleen uitgevraagd onder (ouders van) leerlingen van groep 2. Het is dus niet zo dat gegevens retrospectief worden uitgevraagd (in de zin dat voor leerlingen in groep 8 (aan ouders) wordt gevraagd of ze ooit hebben deelgenomen aan vve).

Niet voor alle scholen in COOL 2010/2011 is informatie beschikbaar over de scores op de Cito-toetsen. In totaal zijn gegevens over Cito-scores beschikbaar voor 318 scholen. Het gaat hierbij om 9.444 leerlingen in de referentiesteekproef en 12.538 in de totale steekproef.

In figuur 5.1 is de gemiddelde score op de Cito-eindtoets weergegeven.



Figuur 5.1 – Gemiddelde score op Cito-eindtoets, groep 8 [Bron: COOL 2010/2011]

Naast de gegevens over de Cito-toets van leerlingen in groep 8 bevat COOL ook gegevens over toetsen van leerlingen in groep 2, 5 en 8.

- Taal voor Kleuters LOVS vaardigheidsscore (groep 2);
- Ordenen vaardigheidsscore (groep 2);
- Woordenschat vaardigheidsscore (groep 5);
- Leeswoordenschat 1999 vaardigheidsscore (groep 8);
- Rekenen/Wiskunde 2002 vaardigheidsscore (groep 8);
- Rekenen/Wiskunde LOVS vaardigheidsscore (groep 5);
- Begrijpend Lezen 1998 vaardigheidsscore (groep 8);
- Begrijpend Lezen LOVS vaardigheidsscore (groep 5);
- DrieMinutenToets 2009 kaart 3 vaardigheidsscore (groep 5).

Programma's vve

In PRIMA en COOL wordt ook vastgelegd of leerlingen op een peuterspeelzaal en/of crèche hebben gezeten en of ze hebben deelgenomen aan een vorm van vve. Voor wat betreft vve wordt onderscheid gemaakt naar type programma. Voor leerlingen die deelnemen aan vve wordt uitgevraagd hoe lang ze hebben deelgenomen. De vragen over peuterzaal, crèche en vve zijn niet retrospectief in de zin dat over leerlingen in groep 8 wordt gevraagd of ze in het verleden hebben deelgenomen.

In tabel 5.1 is voor COOL 2010/2011 weergegeven hoeveel scholen deelnemen aan verschillende onderwijsinterventies, waaronder vve-programma's. Het gaat in een beperkt aantal gevallen om specifieke vve-programma's, namelijk Piramide, Kaleidoscoop en KO-totaal.

Tabel 5.1 - Deelname van scholen aan diverse vve-programma's [Bron: COOL 2010/2011]

Programma	Nee	Ja	In voorbereiding
Voorschool (deelname aan een Voorschoolproject)	217	110	21
Peutergroep binnen de school	251	74	25
Samenwerking met peuterspeelzalen en/of kinderdagverblijven voor 0–4 jarigen	90	245	53
Vversterk	216	108	21
Brede school	183	126	59
Onderwijstijdverlenging	281	45	11
Schakelklas(sen)	263	68	7
Met Woorden in de Weer	268	55	15
Taalpilot / Taalverbetertraject / Taallijn Vve	171	174	21
Rekenpilot / Rekenverbetertraject	253	67	21
Op Maat traject	308	8	5
Project Versterken Opbrengstgericht Werken	221	85	45
Program kleuters: Piramide	206	96	–
Program kleuters: Kaleidoscoop	235	10	–
Program kleuters: Basisontwikkeling	213	56	–
Program kleuters: KO-totaal	223	40	–
Program kleuters: BAS-project	190	65	–
Program kleuters: ander programma	112	102	–

Onderlinge koppelbaarheid COOL/PRIMA

Om de samenhang te bepalen tussen diverse vve-programma's en leerprestaties moet een relatie worden gelegd tussen deelname aan programma's en leerprestaties. Gegeven dat deelname aan programma's in de eerste schooljaren (groep 1 en 2) plaatsvindt en eventuele effecten daarvan pas later (groep 5 en 8) worden gemeten, is het tijdsaspect van belang. Met andere woorden: leerprestaties van leerlingen die in 2010/2011 in groep 5 of 8 zaten moeten worden gerelateerd aan deelname aan programma's toen deze leerlingen in groep 2 zaten. Dit kan niet op basis van afzonderlijke cohorten van PRIMA of COOL omdat gegevens over deelname aan programma's niet retrospectief worden uitgevraagd; van leerlingen die in COOL 2010/2011 in groep 8 zaten, is – op basis van een afzonderlijk cohort – niet bekend of ze in groep 1 en/of 2 hebben deelgenomen aan vve-programma's. Het kan niet worden uitgesloten dat scholen die in 2010/2011 wel deelnamen aan vve-programma's dit voorheen niet deden.

De cohorten van PRIMA en COOL kunnen onderling worden gekoppeld. Dit kan zowel op het niveau van de school als op het niveau van de leerling. Door koppeling van de laatste meting van PRIMA 2002/2003 met COOL 2007/2008 en COOL 2010/2011 kan voor leerlingen die in 2002/2003 in groep 2 zaten de gehele onderwijsloopbaan primair onderwijs in beeld gebracht worden. Op het niveau van de scholen kan dit uiteraard ook. Voor leerlingen die in 2010/2011 in groep 8 zaten, kan door koppeling van de metingen worden bepaald of scholen/leerlingen deelnamen aan vve-programma's.
De koppeling van de laatste meting van PRIMA en COOL 2007/2008 en COOL 2010/2011 is niet perfect. Dit komt door het feit dat scholen niet meer deelnemen aan een vervolgmeting. Dergelijke uitval van scholen is groot; alleen al de uitval van COOL 2007/2008 naar COOL 2010/2011 bedraagt circa vijftig procent.
Na koppeling van de drie meetmomenten (PRIMA, COOL 2007/2008 en COOL 2010/2011) resulteert een overlap van 228 scholen. Van deze scholen is informatie beschikbaar voor al de drie meetmomenten. Dit betekent niet dat voor deze scholen alle leerlingen over de tijd kunnen worden gevolgd. Van de 16.000 leerlingen die in 2002/2003 in groep 2 zaten, zijn in 2010/2011 (groep 8) van ruim 2.000 leerlingen gegevens beschikbaar over deelname aan de eindtoets Cito. Ook hieruit blijkt dat sprake is van grote uitval.

Mogelijkheden en beperkingen
De cohortonderzoeken PRIMA en COOL bevatten een groot aantal kenmerken van leerlingen, ouders en scholen, waaronder deelname aan diverse vve-programma's en andere specifieke interventies en scores op toetsen die de leerprestaties meten. Gegevens worden uitgevraagd op het niveau van de leerling maar kunnen eenvoudig worden geaggregeerd naar het niveau van de scholen. Op basis van PRIMA/COOL kunnen dus relaties tussen vve en leerprestaties worden onderzocht op zowel het niveau van de leerling als school.
Gegevens in PRIMA/COOL worden niet retrospectief uitgevraagd. Zo is bijvoorbeeld van leerlingen in groep 8 in COOL 2010/2011 niet bekend of ze hebben deelgenomen aan vve. Dit wordt alleen uitgevraagd voor leerlingen in groep 2. Om te meten of deelname aan vve van invloed is op leerprestaties van leerlingen (bijvoorbeeld gemeten door toetsen in groep 5 en/of groep 8), moeten de verschillende metingen van PRIMA/COOL onderling worden gekoppeld. Dit kan zowel op het niveau van de leerling als op het niveau van de school. Zo kan door koppeling van PRIMA 2002/2003 met COOL 2007/2008 en COOL 2010/2011 de complete onderwijsloopbaan (primair onderwijs) van leerlingen die in 2002/2003 in groep 2 zaten in beeld worden gebracht (groep 2, groep 5 en groep 8).
Hierbij moet wel rekening worden gehouden met grote uitval. Veel scholen doen na deelname aan PRIMA of COOL niet mee aan een tweede ronde. Naast uitval van scholen is sprake van uitval van leerlingen omdat ze naar een andere school gaan of doubleren. Uit analysen van een gekoppeld bestand blijkt dat van circa 2.000 leerlingen die in 2002/2003 in groep 2 zaten ook gegevens beschikbaar zijn van 2010/2011 van dezelfde leerlingen in groep 8. Dit probleem zal binnen COOL minder groot zijn, omdat leerlingen die aan $COOL_{5-18}$ deelnemen, in het onderzoek betrokken blijven, óók als ze doubleren, van school veranderen of de overstap naar het voortgezet onderwijs maken. Het oorspronkelijke cohort blijft dus (zoveel mogelijk) intact.
Eventuele koppeling van externe data aan COOL is mogelijk. Daarbij dienen specifieke procedures te worden gevolgd. Het CBS zorgt voor de koppeling van de gegevens aan de onderwijsnummer-bestanden.

5.5.3 DUO
Leerlingen
Op verzoek heeft DUO gegevens geleverd van alle leerlingen primair onderwijs in de inschrijfjaren 2009, 2010 en 2011. In het aangeleverde bestand zijn onder meer gegevens opgenomen over:
- geslacht;
- leeftijd;
- woonplaats;
- herkomst;
- deelname aan voor- en vroegschoolse educatie (inclusief duur);
- leerlinggewicht;
- score eindtoets Cito;

- BRIN school.

De gegevens worden niet retrospectief uitgevraagd. Net als in COOL betekent dit dat voor leerlingen in groep 8 niet bekend is of ze hebben deelgenomen aan vve. Het gevolg is dat op het niveau van de leerling geen relatie kan worden gelegd tussen deelnamen aan vve en de leerprestaties (score eindtoets Cito). Uit navraag blijkt dat DUO geen gegevens over oudere inschrijfjaren kan leveren. Dit is het gevolg van de invoering van het leerlingnummer. Het is dus niet mogelijk om deelname aan vve af te leiden uit oudere inschrijfjaren en deze te koppelen aan de leerprestaties van leerlingen in groep 8 in latere inschrijfjaren.

Of wel analysen kunnen worden gedaan op het niveau van de school hangt af van de vraag of externe informatie beschikbaar is over deelname van scholen aan vve in het verleden. Indien dat het geval is, kan een relatie worden gelegd tussen deelname aan vve (op schoolniveau) in, zeg, 2001 en Cito-scores van leerlingen in groep 8 in 2009. Koppeling is eenvoudig te doen op basis van BRIN-nummer.

Kwaliteit data leerlingen

De door DUO geregistreerde gegevens (BRON DUO) zijn afkomstig van de scholen. De ervaring leert dat deze dataverstrekking niet optimaal is. Weliswaar wordt de geleverde informatie op diverse punten gecontroleerd, maar de scholen bepalen – binnen de bestuurlijke afspraken – welke gegevens worden aangeleverd. Op drie essentiële punten is de registratie van DUO niet adequaat en minder goed bruikbaar.

- de gewichtenleerlingen: bekend is dat hierin fouten zitten;
- Cito-scores zijn onvolledig; als gevolg van procedurele belemmeringen en bestuurlijke afspraken zijn de scholen niet verplicht om de Cito-scores van alle leerlingen aan de inspectie (DUO) te verstrekken. Om die reden wijken de Cito-scores op basis van DUO af van de Cito-scores die de scholen rechtstreeks van Cito ontvangen. Scholen laten in de door hen aangeleverde gegevens vooral de laagste Cito-scores weg (van niet-taalvaardige leerlingen en Praktijkonderwijs-leerlingen) en dat is tamelijk bezwaarlijk in het licht van het onderwerp van onderzoek. Niet alle scholen gebruiken Cito en een deel gebruikt een andere toets naast Cito. In het laatste geval worden regelmatig juist de resultaten van de andere toets en niet de Cito-resultaten aan DUO geleverd;
- deelname aan vve wordt onvolledig geregistreerd; uit een door DUO uitgevoerde vergelijking van de Vve-monitor van Sardes met de registratie van DUO is gebleken dat de registratie in BRON slechts een klein deel van de deelname aan vve bevat. Uit de vergelijking blijkt dat het bereik van vroegschoolse programma's van gewichtenleerlingen voor alle basisscholen in de meest conservatieve schatting op 36%, of 14.354 leerlingen wordt geschat (Vve-monitor). De 1.133 (of 3%) gewichtenleerlingen in BRON blijven daarbij ver achter. Een conservatieve schatting van het bereik van gewichtenpeuters bij voorschoolse programma's komt neer op 76% (op basis van de Vve-monitor), waar de registratie in BRON op 12% uitkomt.

Financiële gegevens en inzet personeel

Financiële gegevens (ontvangen bijdragen en subsidies) zijn beschikbaar op bestuursniveau (bevoegd gezag) voor de jaren 2006-2010. De door de besturen ontvangen bedragen betreffen niet alleen de Rijksbijdragen en -subsidies (deels geoormerkt), maar ook gemeentelijke en overige overheidsbijdragen en -subsidies.

Op instellingsniveau zijn er gegevens over niet-personele baten en lasten, uitgesplitst naar hoofdgroepen (administratie, huisvesting, leermiddelen, meubilair, schoonmaak). Ook ontvangen ouderbijdragen en sponsorbedragen zijn hierin opgenomen. Deze gegevens zijn in ieder geval voor 2010 beschikbaar, maar op korte termijn ook voor 2008 en 2009.

Op schoolniveau zijn ook gegevens beschikbaar over de personele inzet (in fte en personen), uitgesplitst naar drie functiegroepen, geslacht en leeftijd. Deze gegevens zijn over een langere periode beschikbaar (in ieder geval 2007-2011).

Deze bestanden van DUO bevatten dus meer informatie over de (feitelijk) beschikbare middelen en inzet dan puur op basis van de gewichtenregeling kan worden afgeleid. Door koppeling van deze informatie aan bijvoorbeeld COOL zouden in de analysen op basis van de COOL-data mogelijk meer specifieke schoolkenmerken kunnen worden meegenomen. Het gaat dan natuurlijk om het vaststellen

van de inspanningen in de periode dat de leerlingen (die recent een eindtoets hebben gedaan) participeerden in vve-activiteiten, dus in de periode 2000-2004.
Op korte termijn is een bezwaar dat direct beschikbare gegevens van DUO niet ver terug gaan in de tijd, maar dat er mogelijk wel informatie is over eerdere jaren (periode 2000-2005) die via maatwerk wel toegankelijk kan worden gemaakt.
Aangetekend moet worden dat DUO ook bij deze gegevens een voorbehoud maakt ten aanzien van de kwaliteit. In deze gevallen heeft het vooral te maken met onvolledigheid (nonrespons van scholen). De dekkingsgraad is echter over het algemeen toch tamelijk hoog.

5.5.4 Vve-monitor

Sinds 2007 voert Sardes in opdracht van het Ministerie van OCW jaarlijks de Landelijke Monitor vve uit. Met deze monitor wordt bijgehouden in hoeverre de landelijke doelstellingen met betrekking tot het aanbod en bereik van vve worden gerealiseerd. Daarnaast wordt inzicht gegeven in de kwaliteit (overdracht, doorgaande lijn, scholing, inzet extra personeel en ouderbetrokkenheid), de opbrengsten, financiën en ontwikkelingen/trends in de voor- en vroegschoolse educatie.
In 2007 en 2008 zijn alleen de onderwijsachterstandengemeenten bevraagd. Vanaf 2009 worden alle gemeenten benaderd voor informatie over het *voorschoolse* gedeelte van vve en basisscholen over de *vroegschoolse* component. In 2011 zijn alle scholen die gewichtengeld krijgen, gebeld voor deelname aan het onderzoek[1]. Gemeenten en schoolbesturen kregen het verzoek een digitale vragenlijst in te vullen over voor- en vroegschoolse educatie. Alle gemeenten hebben een uitnodiging ontvangen. Alleen besturen van scholen die gewichtengeld ontvangen, kregen een verzoek om mee te werken.
Op deze wijze is voor 2011 een aselecte steekproef van 600 achterstandsscholen samengesteld. Van die scholen is bekend wat ze aan onderwijsachterstanden doen (vroegschoolse educatie, klassenverkleining, taalintensivering, extra formatie in onderbouw, verlenging leertijd, schakelklas, ouderbetrokkenheid e.d.). In 2011 zijn tevens extra vragen opgenomen over de inzet van de onderwijsachterstandenbeleidsmiddelen. Van de scholen is bekend of ze de volgende maatregelen inzetten:

- vve-programma's;
- eigen of bovenschoolse klassen;
- klassenverkleining;
- taalcoördinator;
- inzet extra fte in de vorm bijvoorbeeld een tutor;
- overige activiteiten.

Van belang is verder dat scholen verschillen naar de hoeveelheid interventies die men inzet en de intensiteit van de interventies.

Koppelingen aan andere databestanden

Koppeling van de Vve-monitor aan de DUO-registratie zou het bestand dus verrijken, maar weinig toevoegen voor het meten van effecten, vanwege ontbrekende informatie over (a) de eerdere deelname van leerlingen (uit groep 8) aan vve-activiteiten en (b) de inzet van onderwijsachterstandenbeleids-middelen in de afgelopen jaren.
Koppeling van vve aan COOL heeft op korte termijn geen meerwaarde, omdat de mogelijke effecten van de inspanningen in 2011 pas over enige jaren kunnen worden vastgesteld. Ook wordt de deelname aan vve-programma's al in COOL vastgelegd. Wel is denkbaar dat de registratie van de intensiteit van deelname aan vve-programma's in de Vve-monitor beter of anders wordt geregistreerd dan in COOL, of dat deze op meer of andere scholen betrekking heeft, waardoor over enige jaren de meting in het kader van de Vve-monitor kan worden benut voor effectmeting.

[1] In dit jaar zijn tegelijkertijd twee onderzoeken uitgevoerd, één door Oberon (vanuit het ondersteuningstraject Focus Vroegschool) en één door Sardes. De werkzaamheden zijn op elkaar afgestemd om te vermijden dat scholen twee keer zouden worden benaderd.

5.5.5 Mogelijkheden voor effectmeting – een eerste verkenning

SEOR heeft bij wijze van vingeroefening analysen uitgevoerd op basis van de gekoppelde COOL-PRIMA-bestanden. Dit is vooralsnog het meest geschikte bestand om dergelijke analysen mee uit te voeren. Er zijn diverse modellen geschat om de samenhang tussen deelname aan vve op de Cito- eindtoets te meten. Er is in de gespecificeerde modellen geen samenhang gevonden.

Het is echter te voorbarig om op basis van deze resultaten te concluderen dat er geen samenhang zou zijn tussen deelname aan vve en leerprestaties van leerlingen in het primair onderwijs. De modellen zijn bedoeld als vingeroefening en geven de mogelijkheden van PRIMA/COOL voor het meten van effecten van vve aan. Verbetering van het gebruikte model, modelspecificaties etc. zouden mogelijk betere resultaten kunnen leveren, maar het uitgevoerde literatuuronderzoek maakt duidelijk dat de verwachtingen op dit punt niet hooggespannen moeten zijn.

Een specifiek probleem dat wel degelijk van invloed kan zijn op de uitkomsten is dat van selectiviteit op niet-waargenomen kenmerken. Hoewel in het gekoppelde PRIMA/COOL-bestand kan worden gecorrigeerd voor veel factoren, die mogelijk van invloed zijn op de Cito-eindtoets, blijft de mogelijkheid bestaan dat er niet-waargenomen kenmerken zijn die zowel bepalend zijn voor deelname aan vve als de afhankelijke variabele (Cito-score). Indien sprake is van niet-waargenomen selectiviteit kan het effect van vve op leerprestaties niet zuiver worden geschat.

In theorie kan voor selectiviteit op niet-waargenomen kenmerken worden gecorrigeerd. In de praktijk is dit echter lastig omdat geschikte instrumentvariabelen (variabelen die wel van invloed zijn op deelname aan vve maar niet op de output) ontbreken.

5.5.6 Haalbaarheid fase 1 van voorgesteld onderzoeksdesign

De thans voorgestelde opzet van fase 1 van het onderzoeksdesign (zie § 5.6) is haalbaar, maar rekening moet worden gehouden met de eerder gemaakte kanttekeningen over de kwaliteit van de BRON DUO-data in het bijzonder rond de Cito-scores, deelname aan specifieke interventies (vve) en (in mindere mate) de toegekende gewichten.

Dit probleem zou in principe ondervangen kunnen worden als de volledige Cito-scores aan het DUO-bestand zouden kunnen worden toegevoegd. Het is niet zeker of dit – binnen de bestaande bestuurlijke afspraken – is te realiseren.

Verder moet bij de selectie zoveel mogelijk worden gecorrigeerd voor de verschillen in leerlingenpopulatie en schoolkenmerken (regio, e.d.), voor zover die niet in de gewichten meegenomen worden, maar wel van belang zijn voor de Cito-scores.

De mogelijke invloed van andere factoren is ook van belang bij het onderzoeken van de relatie tussen de omvang van de financiële middelen (onderwijsachterstandenbekostiging) en de Cito-scores. Bovendien moet dan rekening worden gehouden met het vertraagde effect van de inspanningen op onderwijsachterstandengebied op de eindtoets.

Zowel de COOL-bestanden, als de DUO-gegevens bevatten in principe voldoende gegevens om een selectie te maken van relevante scholen voor de voorgestelde onderzoeksdesign (verschillend in leerprestaties en aanpak).

5.5.7 Conclusies mogelijkheden bestaande databestanden

De cohortonderzoeken PRIMA en COOL bevatten een groot aantal kenmerken van leerlingen, ouders en scholen, waaronder deelname aan diverse vve-programma's en andere specifieke interventies en scores op toetsen die de leerprestaties meten. Gegevens worden uitgevraagd op het niveau van de leerling maar kunnen eenvoudig worden geaggregeerd naar het niveau van de scholen. Op basis van PRIMA/COOL kunnen dus relaties tussen vve en leerprestaties worden onderzocht op zowel het niveau van de leerling als school.

De verkennende analysen die zijn gemaakt op basis van de gekoppelde PRIMA- en COOL-bestanden bevestigen echter de conclusies uit de literatuurstudie, namelijk dat een verband tussen specifieke inspanningen op het gebied van onderwijsachterstanden en leerprestaties niet eenvoudig is aan te tonen. Dit hangt samen met drie onderling samenhangende factoren, te weten de volledigheid en kwaliteit van de beschikbare data, de meetmethode en de complexiteit van de problematiek.

Het basisidee achter cohortenonderzoek is dat de hele schoolloopbaan van leerlingen in kaart wordt gebracht. Door uitval van scholen en leerlingen neemt echter ook in cohortenonderzoek in de loop van de tijd het aantal waarnemingen waarin de gehele schoolloopbaan wordt bestreken af. Omdat gegevens

in PRIMA/COOL ook niet retrospectief worden uitgevraagd, wordt dit probleem binnen het onderzoek niet opgelost. Wel wordt uitval van leerlingen als gevolg van doublures of verhuizing in $COOL_{5-18}$ beter ondervangen dan in het verleden, doordat deze leerlingen in het onderzoek betrokken blijven.
In de verkennende analysen – bedoeld als vingeroefening – zijn slechts enkele modelvarianten doorgerekend. Verbetering van het gebruikte model, modelspecificaties, etc. zouden mogelijk betere resultaten kunnen leveren, maar het uitgevoerde literatuuronderzoek maakt duidelijk dat de verwachtingen op dit punt niet hooggespannen moeten zijn.
Een specifiek probleem dat wel degelijk van invloed kan zijn op de uitkomsten is dat van selectiviteit op niet waargenomen kenmerken. Hoewel de PRIMA/COOL- bestanden veel gegevens bevatten en er gecorrigeerd kan worden voor veel factoren, blijft de mogelijkheid bestaan dat er niet waargenomen kenmerken zijn die (mede) bepalend zijn voor zowel deelname aan vve als de output variabele (Cito-score).
Het COOL-bestand zou kunnen worden verrijkt door toevoeging van informatie op schoolniveau die voor voorgaande jaren (zeg 2000-2005) in principe beschikbaar is. Deze informatie moet dan nog wel worden bewerkt en er bestaat geen zekerheid dat dan wel een model zou kunnen worden geconstrueerd dat logische samenhangen aantoont. Wel is duidelijk dat wanneer de huidige dataregistratie (zowel via DUO als bijvoorbeeld de Vve-monitor) de komende jaren wordt voortgezet, dat er over enige jaren direct beschikbare informatie aan elkaar kan worden gekoppeld, waardoor een rijk gegevensbestand ontstaat, waarmee naar verwachting betere mogelijkheden voor effectiviteitonderzoek ontstaan. Op korte termijn ligt een meer gericht onderzoeksdesign meer voor de hand.
De besproken databestanden (COOL en DUO) bevatten in principe voldoende informatie over onderwijsinspanningen en (Cito-)resultaten om een gerichte selectie van goed en minder goed presterende scholen te maken.

5.6 Synthese: uitwerking van een onderzoeksdesign

Effectieve interventies voor bestrijding van onderwijsachterstanden: wat werkt?
Op zoek naar goede voorbeelden
Bij de onderzoeksvraag over wat werkt in onderwijsachterstandenbeleid past een combinatie van kwalitatieve en kwantitatieve onderzoeksmethoden. Van goed presterende scholen met veel gewichtenleerlingen kunnen we leren waarom hun programma's succesvol zijn. Door middel van case-studies en kwalitatief onderzoek gaan we op zoek naar '*lessons-to-learn*'. Deze lessen worden in een vervolgtraject ingezet om slecht presterende scholen op het goede spoor te brengen door middel van verbeterprojecten, zodat interventies op een juiste wijze binnen de school inhoud krijgen. De 'harde' resultaten op scholen worden door middel van kwantitatief onderzoek geëvalueerd aan de hand van een aantal indicatoren, waaronder de leerprestaties van de leerlingen.

In de wetenschappelijke onderzoeksliteratuur zijn aanwijzingen te vinden voor het bestaan van effectieve interventies om onderwijsachterstanden te bestrijden op scholen, maar regelmatig kunnen dergelijke effecten niet teruggevonden worden als de onderzoeken worden herhaald. Ook bestaat er onderzoek waarbij de werkzaamheid van interventies niet kan worden aangetoond (Nap-Kolhoff *et al*., 2008; Van Schooten & Sleegers, 2009).
De kwaliteit en de intensiteit van de interventies is voor de effectiviteit van groot belang. Zo zijn in de VS de programma's om onderwijsachterstanden bij peuters te bestrijden zeer intensief en langdurig. De programma's zijn echter nauwelijks vergelijkbaar met het Nederlandse aanbod. Van een fulltime aanbod is niet hetzelfde te verwachten als van een aanbod van maximaal vier dagdelen per week. Dat zou een verklaring kunnen zijn waarom in een aantal Nederlandse onderzoeken de effectiviteit van interventies om achterstanden te bestrijden niet kan worden aangetoond dan wel dat het effect zeer klein is. Vaak gaat het hierbij om onderzoek met een grootschalige kwantitatieve dataverzameling, waarbij de uitvoering van de interventies buiten het zicht zijn gebleven, en zijn de gegevens louter kwantitatief van aard. Veel auteurs in Nederland en in het buitenland menen dat er sprake is van een middelgroot effect. Dit effect is alleen te vinden als de interventie goed wordt uitgevoerd. Leseman (2007) stelt dat in een evaluatiedesign zowel de kwaliteit van de organisatie van belang is, als de inhoud en de continuïteit van de interventie. Omdat de uitvoering van de interventie en de inpassing van de

interventie binnen de school belangrijk is voor het succes, is het van belang om niet alleen in een kwantitatief onderzoeksdesign te kijken naar de leerprestaties van leerlingen uitgedrukt in Cito-scores en deze naast het interventieaanbod te leggen (er is een aanbod: ja of nee) en te registreren hoeveel dagdelen/uren en hoeveel leerlingen hiervan deelnemen profiteren. Dergelijke onderzoeken leveren teleurstellende resultaten op en de werkzame ingrediënten blijven buiten beeld.
Als scholen goede resultaten bereiken met programma's of interventies om onderwijsachterstanden te bestrijden, is het belangrijk om te onderzoeken hoe zij het programma of de interventie gestalte geven en hoe zij het geïmplementeerd hebben. We stellen daarom een gecombineerd kwalitatief en kwantitatief onderzoeksdesign in drie delen voor.

Fase 1: inventarisatie (duur: ½ jaar)
Inventariseren van aantallen goed en slecht presterende scholen met of zonder onderwijsachterstandenbekostiging. Op basis van bestaande data (COOL, DUO, Cito) wordt gezocht naar verschillende scholen, zowel met als zonder onderwijsachterstandenbekostiging, met een groot aantal gewichtenleerlingen (zie voor mogelijkheden van gebruik bestaande databestanden § 5.5).

Gewichtenleerlingen
Scholen krijgen extra geld op basis van het schoolgewicht. De factor 0,06 * aantal leerlingen is een drempel, die ervoor zorgt dat scholen pas extra geld krijgen wanneer meer dan 6 procent van de leerlingen laag of zeer laag opgeleide ouders heeft. Tweederde van alle 7.000 basisscholen in Nederland heeft geen gewicht.
Het bestand waarover het CBS beschikt, bestaat uit gegevens van 1,2 miljoen leerlingen. Het aantal leerlingen met een gewicht is 12%. Niet voor al die leerlingen zijn interventies ingezet. Voor bijna 42.000 leerlingen (=3.5%) geldt dat ze weliswaar een gewicht hebben, maar door de drempel in de regeling krijgen scholen geen gewicht en dus geen extra geld.
Het schoolgewicht wordt als volgt bepaald: a*0,3 + b*1,2 – 0,06*aantal leerlingen, waarbij *a* staat voor het aantal laag opgeleide ouders en *b* voor het aantal zeer laag opgeleide ouders (alleen basisonderwijs).
Door de schoolgewichten naast de scores van de leerlingen op Cito-toetsen te leggen, kunnen goed en slecht presterende scholen worden aangewezen. Dit zijn zowel scholen met onderwijsachterstandenbekostiging als scholen zonder onderwijsachterstandenbekostiging.
Vanuit het databestand kan ook antwoord worden gegeven op de vraag in hoeverre de omvang van de financiële middelen (onderwijsachterstandenbekostiging) invloed heeft op de effectiviteit (het rendement).

Fase 2: kwalitatieve case-studies (duur: 1 jaar)
In navolging van het succesvolle EPPE-project in Engeland wordt een aantal cases bestudeerd. In de tweede fase worden achttien goed en slecht presterende scholen geselecteerd uit het databestand die als *case-study* dienen. Op kwalitatieve wijze worden deze scholen onderzocht op succes- en faalfactoren.

- 6 scholen (3 binnen en 3 buiten de G4) met gemiddelde tot goede Cito-scores op taal en rekenen;
- 6 scholen (3 binnen en 3 buiten de G4) met benedengemiddelde Cito-scores op taal en rekenen;
- 3 scholen zonder onderwijsachterstandenbekostiging met ongeveer 5% gewichtenleerlingen, met gemiddelde tot goede Cito-scores op taal en rekenen;
- 3 scholen zonder onderwijsachterstandenbekostiging met ongeveer 5% gewichtenleerlingen, met benedengemiddelde Cito-scores op taal en rekenen.

Van de scholen met onderwijsachterstandenbekostiging wordt in kaart gebracht op welke wijze de achterstandsmiddelen zijn besteed en voor welke activiteiten die zijn ingezet.
De achttien scholen worden bestudeerd door middel van observatie en checklists en interviews worden afgenomen aangaande de vraag wat de werkzame ingrediënten zijn van de goed presterende scholen en hoe de programma's en hoe interventies worden uitgevoerd. Waarin verschillen de goed presterende scholen van laag presterende scholen en waarin verschilt de uitvoering van het programma of interventie? Er wordt gelet op kwaliteit, intensiteit, instroom leerlingen, *monitoring* van de leerlingen en gebruik van meetinstrumenten en verder de kenmerken bekend uit de (internationale) literatuur. Deze

kenmerken staan in de schema's in de hoofdstukken 2, 3 en 4 genoemd. De drie schema's worden in de samenvatting in elkaar geschoven.
Door middel van kwalitatief onderzoek kunnen 'witte vlekken' met betrekking tot het succesfactoren aan het licht komen. Van succesvolle scholen kunnen we leren waarom hun programma's succesvol zijn. De succesfactoren zijn de input voor een vervolgproject waar laag presterende scholen mee geholpen worden. Ook faalfactoren zijn belangrijk: wat kunnen scholen beter niet doen, wat is niet kosteneffectief, wat levert weinig op?
In het derde deel worden deze '*lessons-to-learn*' meegenomen in een vervolgonderzoek.

Fase 3: Vervolgonderzoek: implementeren van succes in verbeterprojecten (duur: 2½ jaar)
De succesfactoren uit de tweede fase van het onderzoek kunnen worden ingezet in een vervolgonderzoek en een verbeterproject op een aantal laag presterende basisscholen waarbij op prospectieve wijze gegevens worden verzameld aan de hand van een aantal indicatoren (waaronder leerlingprestaties op de Cito-toetsen). De leerlingprestaties van de laag presterende scholen worden tevens in een quasi-experimenteel design vergeleken met laag presterende scholen met een vergelijkbare leerlingpopulatie, die niet aan de verbeterprojecten deelnemen om interventies beter te implementeren en uit te voeren.
Voor dit onderdeel wordt op de negen laag presterende basisscholen uit de tweede fase van het onderzoek een verbeterproject gestart waarbij gebruik wordt gemaakt van de *PDCA-cyclus* (Plan, Do, Check, Act) zoals ook in de verbeterprojecten in de gezondheidszorg of organisatiemanagement gebruikelijk is. Met deze cyclus, waarbij teams aan de slag gaan om verbeteringen door te voeren, ook wel de 'Doorbraakmethode' genoemd, kunnen *good practices* goed binnen een organisatie verspreid worden. De schoolteams worden begeleid om een verbeterplan op te stellen en kunnen direct hiermee aan de slag gaan. Het zwaartepunt van het verbeterplan ligt op een betere kwaliteit en inbedding en uitvoering van stimulerende interventies waarvoor al dan niet onderwijsachterstandenbekostiging is gekregen. Steeds vindt bijstelling plaats volgens de *PDCA-cyclus*. De scholen krijgen een coach en er worden teammiddagen gehouden waarbij de teams aan elkaar kunnen laten zien welke verbeteringen worden doorgevoerd. Afgevaardigden van de goed presterende scholen kunnen hierbij als 'model' of als ambassadeur fungeren en zijn ook bij de teammiddagen aanwezig. Onderzoekers stellen met het team indicatoren op die de verbeteringen in de praktijk meten en waarvan verwacht wordt dat zij ook meetbare resultaten opleveren voor de leerprestaties van de leerlingen. Bij de start van het verbetertraject wordt een nulmeting gedaan onder de leerkrachten / leidsters (vragenlijsten en observatie), schoolleiders (gestructureerd interview) en leerlingen (Cito-gegevens, Tak-scores).
Na een jaar en na een vervolgjaar worden gegevens verzameld. Er zijn drie meetmomenten en de gehele dataverzamelingsperiode duurt twee jaar. Verwachting is dat scholen die aan een verbeterproject deelnemen, betere leeropbrengsten kunnen laten zien en beter gestalte kunnen geven aan de interventies om onderwijsachterstanden te bestrijden. De Cito-gegevens van de scholen die deelnemen aan het verbetertraject, kunnen worden vergeleken met een controlegroep van laag presterende scholen met een vergelijkbare populatie gewichtenleerlingen. Er kan bij de start gematcht worden op wijkkenmerken, leerlinggewichten en Cito-gegevens. Met betrekking tot de scholen uit de controlegroep kan er gebruik gemaakt worden van bestaande datasets zodat de scholen uit de controlegroep niet belast worden met gegevensverzameling. Als taal- en rekenprestaties na twee jaar tussen studie- en controlegroepen gaan verschillen is het redelijk om te veronderstellen dat de school de verschillen verklaart.

6 Beschouwing

In dit afsluitende hoofdstuk worden enkele van de hiervoor gepresenteerde onderzoeksresultaten nog eens beschouwd. Bij sommige bevindingen worden kanttekeningen geplaatst, voor andere resultaten worden mogelijke verklaringen aangedragen.

Wat het meest opvalt in onze literatuurstudie, is dat het bewijs voor de effectiviteit van bepaalde interventies ter bestrijding van onderwijsachterstanden mager is. Voor een belangrijk deel berust dit op de gebruikte onderzoeksdesigns en verder ook op de onderzoeksresultaten.
Om met de onderzoeksdesigns te beginnen, wordt teruggegrepen op § 5.3 waar overwegingen gepresenteerd worden voor de keuze van een onderzoeksdesign. Eerder zijn in § 5.2 vanuit de literatuur allerlei op- en aanmerkingen geplaatst over methodologische beperkingen van onderzoek naar onderwijsachterstanden. Deze hebben te maken met de aard van het onderzoek, de onderzoekspopulatie en verstorende variabelen en verder met de interventies zelf, de gekozen uitkomstmaat en/of de setting waarin het onderzoek vorm kreeg. Met die kanttekeningen in het achterhoofd is in § 5.3 geschetst waaraan een onderzoeksontwerp, met behulp waarvan effectiviteit van een interventie onderzocht wordt – in dit geval op het gebied van onderwijsachterstanden – idealiter moet voldoen.
Ongetwijfeld hebben vele onderzoekers wier bevindingen ten grondslag liggen aan deze literatuurstudie, dezelfde overwegingen gehad bij het opstellen van hun design. Het theoretisch ideaaltype voor effectiviteitsonderzoek is in de praktijk veelal niet haalbaar. De belangrijkste drempel lijkt toch wel een financiële te zijn. Gedegen onderzoek, waarmee in dit kader wordt bedoeld: onderzoek met een design dat in staat stelt causale verbanden te leggen en daardoor het onderzoek robuuste bewijskracht verleent, is nu eenmaal kostbaar. Dit heeft enerzijds met de schaal van een dergelijk onderzoek te maken; hierbij valt niet alleen te denken aan de grootte van de onderzoekspopulatie (met als aandachtspunten representativiteit en extrapoleerbaarheid), maar ook aan allerlei te valideren meetinstrumenten voor de afhankelijke en onafhankelijke variabelen en – gebleken niet onbelangrijk – covariaten, waarvoor in analysen gecorrigeerd moet worden om het vaststellen van zuivere effecten mogelijk te maken. Anderzijds is gedegen onderzoek duur vanwege de tijd die ermee gemoeid is, zeker in longitudinale studies.
Naast deze randvoorwaardelijke aspecten kunnen ook methodologische en ethische argumenten aangevoerd worden om voor een minder sterk design te kiezen. Zo kan het in de praktijk onmogelijk zijn om de interventie aan slechts een deel van de leerlingen toe te wijzen. Het kan in bepaalde gevallen ook als onethisch ervaren worden bepaalde (etnische) groepen af te zonderen als onderzoekssubject of anderen een veelbelovende interventie te onthouden.
Al met al gaat het erom een optimaal onderzoeksdesign te ontwikkelen waarin, binnen de beperkingen van bijvoorbeeld geld, tijd en onderzoekspopulatie, de maximale sterkte van een design gerealiseerd kan worden.
Onderzoeksdesign B (zie § 5.4) is methodologisch gezien duidelijk sterker dan de beide andere voorgestelde onderzoeksopzetten. In een voorbespreking met de opdrachtgever bleek de voorkeur echter uit te gaan naar het pragmatisch gezien meer haalbare design C. Het moet gezegd: hiermee schaart het voorgestelde onderzoek zich in de rij van onderzoeken die niet aan de gouden standaard voldoen. In opzet C worden voor de Nederlandse situatie faal- en succesfactoren van interventies geïnventariseerd; het praktijkdeel (de implementatie van succesvolle interventies) is zo ingericht dat onder de omstandigheden (kleinschalig) het beste bewijs kan worden verkregen.

Los van soms inadequate onderzoeksdesigns, tenminste om effectiviteit aan te tonen, speelt dat de onderzoeksresultaten ook niet allemaal eensluidend zijn. Omdat hoofdstuk 2 helemaal in het teken staat van vve, een interventie die an sich bedoeld is ter voorkoming en bestrijding van onderwijsachterstanden, wordt dit als voorbeeld genomen, maar het komt ook wel met betrekking tot andere interventies voor. Heel duidelijk kan het onderzoek gegroepeerd worden naar aangetoonde effecten enerzijds en niet-aangetoonde effecten anderzijds. Het zit ‘m voornamelijk in de kwaliteit van

de vve-activiteit, want daarin blijkt qua intensiteit (aantal maanden en aantal keren per week), ouderbetrokkenheid en personeelskwaliteit nogal wat verscheidenheid te zijn.
Dat vele onderzoeken een wisselend beeld laten zien, onderkennen ook Nap-Kolhoff *et al*. (2008). Zij noemen als verklaring voor het uitblijven van resultaten dat de wijze van implementatie van vve vaak niet is meegenomen in het onderzoek en er geen zicht is op goede uitvoeringscondities. De auteurs adviseren om diepgaand en wetenschappelijk onderzoek te doen naar de effecten van vve zoals een experimenteel onderzoek met een nulmeting en een controlegroep waarbij kwalitatief rijk materiaal wordt verzameld over de leerlingen en over de manier waarop het programma of de interventie wordt uitgevoerd.
Driessen (2012) stelt in een recent document dat van de meeste erkende programma's slechts in zeer beperkte mate onderzoek is gedaan naar de effectiviteit, en voor zover daarvan al sprake is, van geen enkel programma is aangetoond dat het echt werkt, hooguit op onderdelen en onder optimale condities. Geen enkel programma komt dan ook verder dan het laagste officiële erkenningsniveau 'theoretisch goed onderbouwd'. Hierbij geldt dat alleen de oorspronkelijke programma's zijn beoordeeld; over de effectiviteit van lokale aanpassingen daarvan en lokaal ontwikkelde programma's is helemaal niets bekend. Het ontbreken van informatie over betrouwbaarheid, validiteit en effectiviteit van de bij vve gebruikte instrumenten en programma's vormt een serieus manco voor het functioneren van en het realiseren van resultaten met vve.
Van instrumenten die gemeenten gebruiken voor de signalering, screening, indicatiestelling, *monitoring*, werving, toeleiding en overdracht is van praktisch geen enkel instrument onderzocht of het voldoet aan de gangbare eisen van betrouwbaarheid en validiteit, en in de weinige gevallen als dat wel is gebeurd, luidt de conclusie dat dit niet het geval is.
Al eerder is geconstateerd (§ 5.2) dat de setting waarin buitenlands onderzoek verricht werd, veelal niet vergelijkbaar is met de Nederlandse situatie en onderzoeksresultaten daardoor minder relevant zijn. Volgens Kassenberg (2010) is er vooral veel Amerikaans onderzoek, omdat men daar al een lange traditie heeft van verlenging van de schooltijd. Ook de wet '*No Child Left Behind*' heeft bijgedragen aan de aandacht van onderzoekers voor onderwijsachterstanden.
Muijs *et al*. (2004) bevelen aan meer te theoretiseren over bestrijding van onderwijsachterstanden, zonder daarbij te willen komen tot een paraplumodel van school(prestatie)verbetering, omdat een dergelijk model te weinig rekening houdt met de schoolspecifieke context. Niet voor niets wordt 'goed' onderzoek gerepliceerd bij andere doelgroepen en onder andere omstandigheden, alvorens van 'algemeen geldende' effectiviteit of werkzaamheid sprake kan zijn.

Graag hadden de onderzoekers een soort rangordening aangebracht in de beschreven onderzoeken, waarbij experimenten een grote zeggingskracht zouden krijgen en minder robuust (descriptief, correlationeel of kwalitatief) onderzoek meer als aanvullend bewijs beschouwd zouden worden. Deze opzet bleek in de praktijk niet mogelijk. Gezien de korte doorlooptijd van dit project (drie maanden) is in eerste instantie gebruik gemaakt van *narrative* en systematische reviews. Hierin vatten auteurs tal van ter zake doende artikelen samen, in woorden dan wel herberekende effectmaten. De onderliggende artikelen zijn soms experimenten (zeker bij meta-analysen), maar in de verhalende reviews wordt ook wel kwalitatief onderzoek betrokken. Hierdoor loopt de bewijskracht van het basismateriaal door elkaar. Patall *et al*. (2010) constateren ook dat eerdere reviews over schooltijdverlenging slechts tentatief bewijs opleveren. Zij raden aan zich te richten op de '*instructional quality*' en effectief gebruik van de schooltijd (meet de tijd die extra besteed wordt aan prestatieverbetering (*academic achievement*) en niet zonder meer de extra schooltijd *an sich*; dan kan er een sterker verband gevonden worden (groter contrast). Even later stellen zij overwogen te hebben een meta-analyse uit te voeren, maar daarvan hebben afgezien vanwege de heterogeniteit van de gebruikte onderzoeksdesigns (Patall *et al*., 2010).
Meta-analysen zijn geselecteerd op publicatiedatum en gaan vaak verder terug in de tijd. Het komt geregeld voor dat studies van na 1985 in de meta-analysen worden geïncludeerd. In dat jaartal verscheen een studie van Karweit (1985), dat voor veel auteurs blijkbaar als standaardwerk geldt.

De onderzoekers van dit project hebben een beredeneerde keuze gemaakt om zich aanvankelijk alleen te richten op reviews van verschillende aard. Deze keuze is vooral ingegeven door de korte doorlooptijd en het grote terrein dat het literatuur onderzoek in principe bestrijkt. Uitgangspunt was wat de literatuur te zeggen heeft over effectieve interventies op dit gebied. Dat had wel als consequentie dat

specifieke interventies niet als zoekterm zijn opgevoerd in de searches. Daardoor is er ook wel literatuur gemist, zoals onderzoeksliteratuur over klassenverkleining of de insteek van Dronkers.
Er is duidelijk gezocht naar literatuur die over het verbeteren van schoolprestaties gaat in geval van onderwijsachterstanden. Als een artikel over het bevorderen van schoolprestaties van een algemene populatie leerlingen ging, is het terzijde geschoven. Het blijkt dat onderwijsachterstandenbeleid toch wel een heel Nederlands begrip is. Om toch voldoende literatuur te kunnen verwerken, is ervoor gekozen literatuur te verzamelen die gaat over groepen die veelal met onderwijsachterstanden te maken hebben. Vaak betreft dit kinderen van laagopgeleide ouders (meer in het algemeen: lage sociaaleconomische status: *low-SES*) en kinderen van ouders met een allochtone achtergrond. De definitie van onderwijsachterstandenbeleid van Ledoux *et al*. (2011) wijst al hierop. Daarmee is niet gezegd dat etniciteit synoniem is voor onderwijsachterstand; er zijn ook kinderen die aan dit criterium voldoen en prima meekomen op school en andersom: kinderen met een autochtone achtergrond die juist wel een achterstand oplopen.

Bibliografie

Algera, M., Bulk, L. van den & Schooten, E. van (2011). *De Children's Zone op Zuid: effecten van meer leertijd op taal en rekenen, effecten van meer leertijd op soft skills*. Rotterdam: Kenniswerkplaats Rotterdams Talent.

Anderson, L.M., Shinn, C.M.S., Fullilove, M.T., Scrimshaw, S.C., Fielding, J.E., Normand, J. & Carande-Kulis, V.G. (2003). The effectiveness of Early Childhood Development Programs; a systematic review. *American Journal of Preventive Medicine* 24; p.32-46.

Appelhof, P. (2009). *Een oriëntatie naar verlengde onderwijstijd; inrichting en effecten*. Utrecht: Oberon.

Beekhoven, S., Jepma, IJ. & Kooiman, P. (2010). *Landelijke Monitor Voor- en Vroegschoolse Educatie 2010; de vierde meting*. Utrecht: Sardes.

Biedinger, N. (2011). The influence of education and home environment on the cognitive outcomes of preschool children in Germany. *Child Development Research*, doi:10.1155/2011/916303; p. 1-10.

Bijl, R., Boelhouwer, J., Cloïn, M. & Pommer, E. (Red.) (2011). *De sociale staat van Nederland 2011*. Den Haag: Sociaal en Cultureel Planbureau.

Blok, H., Fukkink, R.G., Gebhardt, E.C. & Leseman P.P.M. (2005). The relevance of delivery mode and other programs characteristics for the effectiveness of early childhood intervention with disadvantaged children. *International Journal of Behavioral Development* 29 (1); p. 35-47.

Bulk, L. van den & Algera, M. Effecten van meer leertijd op niet-cognitieve vaardigheden. *In*: Algera, M., Bulk, L. van den & Schooten, E. van (2011). *De Children's Zone op Zuid: effecten van meer leertijd op taal en rekenen, effecten van meer leertijd op soft skills*. Rotterdam: Kenniswerkplaats Rotterdams Talent.

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood research Quarterly* 25 (2); p.140-165.

Burgess, M. & Dixon, P. (2012). A comparative study of school effectiveness in Aguablanca, Colombia.*Educational Management Administration & Leadership*40; p. 362-375.

Clark, R., Harris, A. & Allen, W. (2005). After-school programs: how they affect black male development and educational progress. *Challenge* 11 (2); p. 1-37.

Clinton, J., Hattie, J.A.C. & Dixon, R. (2007). *Evaluation of the Flaxmere Project: When families learn the language of school*. Wellington, New Zealand: Ministry of Education, New Zealand.

Dagevos, J., Gijsberts, M. & Praag, C. van (red.) (2003). *Rapportage minderheden: Onderwijs, arbeid en sociale integratie*. Den Haag: Sociaal en Cultureel Planbureau.

Driessen, G.W.J.M. (2004). A large-scale longitudinal study of utilization and effects of early childhoodeducation and care in the Netherlands. *Early Child Development and Care* 174 (7-8); p.667-689.

Driessen, G. (2012). *Variatie in voor- en vroegschoolse educatie,een onderzoek naar de uiteenlopende wijzen waarop in de gemeenten vorm wordt gegeven aan vve*. Nijmegen: ITS, Radboud Universiteit Nijmegen.

EPPE (2004). *The effective provision of preschool education (EPPE) project (Final Report)*. London: University.

Ewijk, R. van & Sleegers, P. (2010). Peer ethnicity and achievement: a meta-analysis into the compositional effect. *School effectiveness and school improvement: an international journal of research, policy and practice* 21 (3); p.237-265.

Faubert, B. (2012). A literature review of school practices to overcome school failure. *OECD Education Working papers*, no. 68. OECD Publishing.

Geddes, R., Frank, J. & Haw, S. (2011). A rapid review of key strategies to improve the cognitive and social development of children in Scotland. *Health Policy* 101; p. 20-28.

Gevers Deynoot-Schaub, M.J.J.M., Fukkink, R.G., Riksen-Walraven, J.M.A., De Kruif, R.E.L., Helmerhorst, K.O.W., & Tavecchio, L.W.C. (2009). *De NCKO-Kwaliteitsmonitor*. Amsterdam: Nederlands Consortium Kinderopvang Onderzoek.

Gorey, K.M. (2001). Early childhood education: a meta-analytic affirmation of the short- and long-term benefits of educational opportunity. *School PsychologyQuarterly* 16; p. 9-30.

Groot, W., Maassen-van den Brink, H., Dobbelsteen, S. & Mierlo, N. van (2001). *The economics of early childhood education*. Paris: OECD.

Harrison, L.J., Goldfield, S., Metcalfe, E. & Moore, T. (2012). *Early learning programs that promote children's developmental and educational outcomes*. Resource sheet no.15 produced for the Closing the Gap Clearinghouse.

Hattie, J. (2009). *Visible learning; a synthesis of over 800 meta-analyses relating to achievement*. London/New York: Routledge.

Hattie, J. (2012). *Visible learning for teachers; maximizing impact on learning*. London/New York: Routledge.

Heckman, J.J. (2000). Policies to foster human capital. *Research in Economics* 54; p.3-5682.

Herweijer, L. & Turkenburg, M. (2012). *Op zoek naar bewijs; evaluatieontwerpen onderwijs-maatregelen*. Den Haag: Sociaal en Cultureel Planbureau.

Inspectie van het Onderwijs (2008). *De kwaliteit van voor- en vroegschoolse educatie in de vier grote steden; pilot: toezicht op voor- en vroegschoolse educatie in de G4*. Utrecht: Inspectie van het Onderwijs.

Isaacs, J. (2008). *Impacts of early childhood program*. Washington: First Focus, Center on children & families at the Brookings institution.

Jeynes, W.H. (2007). The relationship between parental involvement and urban secondary school and student achievement: a meta-analysis. *Urban Education* 42 (1); p.82-110.

Karweit, N. (1985). Should we lengthen the school term? *Educational Researcher* 14 (6); p. 9-15.

Kassenberg, A. (2010). *Literatuurstudie Uitgebreid Onderwijs*. Groningen: Hanze-hogeschool Groningen, lectoraat Integraal Jeugdbeleid.

Kruijf, R.E.L. de, Riksen-Walraven, J.M.A., Gevers Deynoot-Schaub, M.J.J.M., Helmerhorst, K.O.W., Tavecchio, L.W.C., & Fukkink, R.G. (2009). *Pedagogische kwaliteit van de opvang voor 0- tot 4-jarigen in Nederlandse kinderdagverblijven in 2008*. Amsterdam/Nijmegen: Nederlands Consortium Kinderopvang Onderzoek.

Lauer, P.A., Akiba, M., Wilkerson, S.B., Apthorp, H.S., Snow, D. & Martin-Glenn, M.L. (2006). Out-of-school-time programs: a meta-analysis of effects for at-risk students. *Review of educational research* 76; p. 275-313.

Ledoux, G., Roeleveld, J., Driessen, G., Cuppen, J. & Meijer, J. (2011). *Prestaties en loopbanen van doelgroepleerlingen in het onderwijsachterstandenbeleid; stand van zaken en ontwikkelingen in de periode 1994-2007.* Amsterdam / Nijmegen: Kohnstamm Instituut / ITS.

Leseman, P.P.M. (2002). *Onderzoek in de voor- en vroegschoolse periode: trends en nieuwe aanvragen*. Den Haag: NWO programmaraad Onderwijsonderzoek.

Leseman, P.P.M. (2007). *Early Education for Immigrant Children*. MPI, Bertelsmann Stiftung: The Transatlantic Task Force on Immigration and Integration. Utrecht: Utrecht University.

MacBeath, J. (2011). *Children's University; national evaluation*. Cambridge: University of Cambridge.

Meij, H., Mutsaers, K. & Pennings, T. (2009). *Effectiviteit van voor- en vroegschoolse programma's in Nederland*. Utrecht: Nederlands Jeugdinstituut (NJi).

Meijnen, W. (red.) (2004). *Onderwijsachterstanden; Update van een review. Recente ontwikkelingen inzake onderwijsachterstanden*. Amsterdam: SCO-Kohnstamm Instituut.

Meyer, E. & Klaveren, C. van (2011). *Evaluation of an extended day program in the Netherlands: a randomized field experiment.* Maastricht: TIER.

MO-Groep (2001). *Peuterspeelzaalwerk in de 21ste eeuw*. Utrecht: MO-Groep.

Muijs, D., Harris, A., Chapman, C., Stoll, L. & Russ, J. (2004). Improving schools in socioeconomically disadvantaged areas – a review of research evidence. *School effectiveness and school improvement: an international journal of research, policy and practice* 15 (2); p. 149-175.

Mulder, L., Veen, I. van der, Derriks, M. & Elshof, D. (2012). *De schakelklasleerlingen verder gevolgd. Het tweede vervolgonderzoek bij leerlingen die in 2006/07 of 2007/08 in een schakelklas hebben gezeten*. Nijmegen/Amsterdam: ITS Radboud Universiteit Nijmegen / Kohnstamm Instituut.

Mutsaers, K., Zoon, M., Baat, M. de (2012). *Wat werkt bij onderwijsachterstanden?* Utrecht: Nederlands Jeugdinstituut (NJi).

Nap-Kolhoff, E., Schilt-Mol, T. van, Simons, M., Sontag, L., Steensel, R. van & Vallen, T. (2008). *Vve onder de loep; een studie naar de uitvoering en effectiviteit van voor- en vroegschoolse educatieve programma's*. Tilburg: IVA.

OCW (2000). Regeling Voor- en vroegschoolse Educatie (vve). *Uitleg*, 18a (1); p.54-60.

OCW (2006). Besluit vaststelling doelstelling en bekostiging onderwijsachterstandenbeleid (OAB) 2006-2010. *Staatsblad* 451, 's-Gravenhage: Sdu Uitgevers.

OCW (2010). Besluit specifieke uitkeringen gemeentelijk onderwijsachterstandenbeleid 2011-2014. *Staatsblad* 687, 's-Gravenhage: Sdu Uitgevers.

OECD (2006). *Starting Strong II: Early Childhood Education and Care*. Parijs: OECD.

Onderwijsraad (2002). *Spelenderwijs*. Den Haag: Onderwijsraad.

Onderwijsraad (2011). *Uitgebreid Onderwijs*. Den Haag: Onderwijsraad.

Patall, E.A., Cooper, H. & Batts Allen, A. (2010). Extending the school day or school year: a systematic review of research (1985-2009). *Review of educational research* 80; p. 401-436.

Peisner-Feinberg, E. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child development* 72 (5); p.1534-1553.

Reynolds, A., Magnuson, K. & Ou, S-R (2006). PK-3 Education: programs and practices that work in children's First decade. *Foundation for Child Development*; advancing PK-3, no. 6; p. 1-28.

Reynolds, A.J., Mann, E., Miedel, W. & Smokowski, P. (1997). The state of early childhood intervention: effectiveness, myths and realities, new directions. *Focus* 19 (3); p.25-28.

Reynolds, A.J., Temple, J.A., Ou, S.-R., Arteaga, I.A. & White, B.A.B. (2011). School based early childhood education and age-28 well-being: effects by timing, dosage, and subgroups. *Science* 333 (6040); p.360-364.

Reynolds, A.J., Temple, J.A., Ou, S.-R., Robertso, D.L., Mersky, J.P., Topitzes, J.W. & Niles, M.D.(2007). Effects of a school-based, early childhood intervention on adult health and well being: a 19-year follow up of low-income families. *Archives of Pediatric and Adolescent Medicine* 161; p.730-739.

Roth, J.L., Malone, L.M & Brooks-Gunn, J. (2010). Does the amount of participation in afterschool programs relate to developmental outcomes? A review of the literature. American Journal of Community Psychology, 45, 310-324.

Schooten, E. van (2012). Effecten van meer leertijd op taal en rekenen. *In*: Algera, M., Bulk, L. van den & Schooten, E. van (2011). *De Children's Zone op Zuid: effecten van meer leertijd op taal en rekenen, effecten van meer leertijd op soft skills*. Rotterdam: Kenniswerkplaats Rotterdams Talent.

Schooten, E. van & Sleegers, P. (2009). *Onderzoek naar de effectiviteit van vve- en peuterspeelzalen in Oosterhout en Den Bosch*. Amsterdam: SCO-Kohnstamm Instituut.

Schweinhart, L.J., Montie, J., Xiang, Z., Barnett, W.S., Belfield, C.R. & Nores M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40*. Ypsilanty, Michigan: High/Scope Press.

Silva, E. (2007). *On the clock; rethinking the way schools use time*. Washington DC: Education Sector.

Steensel, R.C.M. van, Kurvers, J.J.H. & Vallen, A.L.M. (2007). De effectiviteit van voorschoolse kindercentra en de invloed van centrumkenmerken en gezinsachtergrond. *Pedagogische Studiën* 84; p.193-206.

Terzian, M., Anderson Moore, K. & Hamilton, K. (2009). *Effective and promising summer learning programs and approaches for economically-disadvantaged children and youth.* Wallace Foundation. White paper.

Veen, A., Fukkink, R. & Roeleveld, J. (2006). *Evaluatie van Startblokken en basisontwikkeling: implementatie en effecten*. Amsterdam: SCO-Kohnstamm Instituut.

Veen, A., Roeleveld, J. & Daalen, M. van (2008). *Implementatie en effecten van voor- en vroegschoolse educatie in Rotterdam; samenvatting*. Amsterdam: SCO-Kohnstamm Instituut, Universiteit van Amsterdam (SCO-rapport nr 803).

Veen, A., Roeleveld, J. & Leseman, P. (2000) *Evaluatie van Kaleidoscoop en Piramide– eindrapportage 2000*. Amsterdam: SCO-Kohnstamm Instituut.

Vegt, A.L. van der (2009). *Voorzieningen voor kinderen van 0-4 jaar: stand van zaken en actuele ontwikkelingen*. Utrecht: Sardes.

Vegt, A.L. van der, Studulski, F. & Kloprogge, J. (2007). *Voorschoolse voorzieningen en onderwijskansen; review van de onderzoeksliteratuur*. Utrecht: Sardes.

Vermeer, H.J., IJzendoorn, M.H., Kruijf, R.E.L. de, Fukkink, R.G., Tavecchio, L.W.C., Riksen-Walraven, J.M.A. & Zijl, J. van (2005). *Kwaliteit van Nederlandse kinderdagverblijven: trends in kwaliteit in de jaren 1995-2005*. Amsterdam: Nederlands Consortium Kinderopvang Onderzoek.

What Works Clearinghouse (2008). *Reading recovery*. Institute of Education Sciences; p. 1-15.

What Works Clearinghouse (2012). *WWC Review of the report 'The effectiveness of a program to accelerate vocabulary development in kindergarten'*. Institute of Education Sciences; p. 1-7.

Wolfe, B., & Tefft, N. (2007). Child interventions that may lead to increased economic growth. *EarlyChildhood Research Collaborative Discussion Paper* 111. Madison, WI: University of Wisconsin-Madison.

Woodland, M.H. (2008). Whatcha doin' after school? A review of the literature on the influence of after-school programs on young black males. *Urban education* 43 (5); p. 537-560.

CED-Groep
Postbus 8639, 3009 AP Rotterdam, bezoekadres: Dwerggras 30
T: + 31 10 4071 599

Locatie Maartensdijk
Postbus 25, 3738 ZL Maartensdijk, bezoekadres: Tolakkerweg 153
T: + 31 346 219 777

E: info@cedgroep.nl