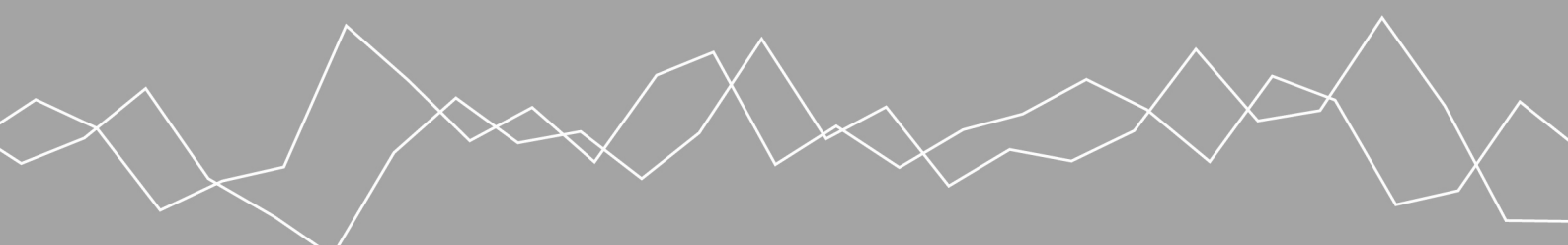


Literatuuronderzoek effectiviteit van onderwijsinterventies



seo economisch onderzoek

Amsterdam, februari 2007
In opdracht van het ministerie van OCW voor de Kenniskamer OCW februari 2007

Literatuuronderzoek effectiviteit van onderwijsinterventies

M.L. Biermans
m.m.v. dr. P. Berkhout

A decorative horizontal line art graphic consisting of several overlapping, jagged, and irregular lines in a light gray color, spanning the width of the page.

seo economisch onderzoek

“De wetenschap dat het goed is”

SEO Economisch Onderzoek doet onafhankelijk toegepast onderzoek in opdracht van overheid en bedrijfsleven. Ons onderzoek helpt onze opdrachtgevers bij het nemen van beslissingen. SEO Economisch Onderzoek is gelieerd aan de Universiteit van Amsterdam. Dat geeft ons zicht op de nieuwste wetenschappelijke methoden. We hebben geen winstoogmerk en investeren continu in het intellectueel kapitaal van de medewerkers via promotietrajecten, het uitbrengen van wetenschappelijke publicaties, kennisnetwerken en congresbezoek.

SEO-rapport nr. 964

ISBN 13 978-90-6733-389-4

Inhoudsopgave

1	Inleiding.....	1
2	Onderwijspersoneel.....	3
3	Voortijdig schoolverlaters / leer-werkplicht.....	11
4	Voor- en vroegschoolse educatie.....	13
5	Schaalgrootte.....	15
6	Examinering.....	17
7	Conclusies en Aanbevelingen.....	19
Literatuur	25
Bijlage	27

1 Inleiding

Onderzoek naar de effecten van de onderwijs wordt steeds vaker onderverdeeld in de traditionele of 'naïeve' evaluaties enerzijds en nieuw of 'hard' onderzoek anderzijds.¹ Het verschil wordt bepaald in hoeverre er in het onderzoek rekening is gehouden met de causaliteit en dus met het probleem van endogeniteit. Dit soort onderzoeken hebben een (quasi-) experimentele opzet met een controlegroep. Het Ministerie van OCW wil inzicht verkrijgen in de meest recente ontwikkelingen op het gebied van wetenschappelijk onderzoek naar onderwijsinterventies waarbij dit probleem expliciet is aangepakt. SEO Economisch Onderzoek is gevraagd om op een vijftal thema's een inventarisatie te maken van de recente publicaties die hieraan voldoen. Dit overzicht geeft tevens aan alwaar de zogenaamde 'witte vlekken' zich bevinden en bij welke thema's geen of bijzonder weinig 'harde' evaluaties voorhanden zijn.

In overleg met het ministerie van OCW is besloten om de literatuurscan te concentreren op de volgende vijf thema's:

- A) Onderwijspersoneel
- B) Voortijdig schoolverlaten / leer-werkplicht
- C) Voor- en vroegschoolse educatie (VVE)
- D) Schaalgrootte (van de scholen/ vestigingen)
- E) Examinering (de wijze van, frequentie, e.d.)

Het eerste thema 'Onderwijspersoneel' is verder uitgesplitst in vijf onderdelen: 1) de relatie tussen de kwaliteit van leraren en de kwaliteit van het onderwijs, 2) de relatie tussen de kwaliteit van leraren en hun beloning, 3) de relatie tussen de kwaliteit van het onderwijs en de beloning van onderwijspersoneel, 4) de mogelijkheden tot verbetering van de kwaliteit van leraren en 5) de rol van ondersteunend onderwijspersoneel. Het mag duidelijk zijn dat de geselecteerde thema's niet op zich staan en dat er vele onderlinge verbanden zijn. Zo is het voortijdig schoolverlaten nauw verbonden met voor- en vroegschoolse educatie.

De tweede factor, naast de reeds genoemde eis van causaliteitsbepaling, die de reikwijdte van de inventarisatie bepaalde was dat het onderzoek enige relevantie voor de Nederlandse situatie dient te hebben. Hiermee wordt niet bedoeld dat het onderzoek per se binnen Nederland moet hebben plaatsgevonden maar dat de uitkomsten ook van betekenis kunnen zijn voor het Nederlandse beleid. Wanneer het gaat om dit type onderzoek staan de Verenigde Staten bovenaan. Deugdelijk onderzoek naar de effecten van onderwijsinterventies is nog relatief schaars. Uit de literatuur komt naar voren dat de institutionele omgeving invloed heeft op de aard en omvang van de uiteindelijke effecten van onderwijsinterventies. Daaruit volgt dat deze effecten dus kunnen afwijken in een andere institutionele setting. De boodschap is dan ook dat de uitkomsten van internationaal onderzoek niet altijd van toepassing zullen zijn voor Nederland.

Het meest belangrijke selectie criterium is echter dat het onderzoek een experimentele of quasi-experimentele opzet moet hebben, oftewel de aanwezigheid van een controlegroep. De opgenomen literatuur kenmerkt zich doordat er in het onderzoek rekening is gehouden met de

¹ Zie bijvoorbeeld CPB (2006), Webbink (2005) en Leigh & Ryan (2005)

aard en de richting van de causaliteit. Webbink (2005) geeft een overzicht van verschillende onderzoeken waarbij de uitkomsten zijn weergegeven zonder dat er rekening wordt gehouden met endogeniteit van de uitkomsten en waarbij dit wel het geval is. Een direct gevolg is dat er een veel eenduidiger beeld naar voren komt waarbij de uitkomsten van verschillende studies elkaar niet meer tegenspreken (zowel positieve als negatieve effecten voor dezelfde interventie). De boodschap is dat het negeren van de endogeniteit kan leiden tot onjuiste conclusies.

In de afgelopen jaren is een drietal overzichtsstudies verschenen met betrekking tot de effectiviteit van onderwijsinterventies waarbij het zwaartepunt lag bij (quasi) experimenteel onderzoek: Webbink (2005), van der Steeg (2005) en Cornet, Huijzinga, Minne en Webbink (2006). Deze drie publicaties geven samen een uitgebreid en nauwgezet overzicht van de relevante literatuur over de effectiviteit van onderwijsinterventies. Het doel van dit literatuuroverzicht is om zoveel mogelijk aan te sluiten bij deze publicaties en daarmee te fungeren als een soort *update*². Eén verdere opmerking moet hier nog worden gemaakt ten aanzien van het type onderzoek dat in dit overzicht is opgenomen. Hoewel het, zoals aangegeven, de bedoeling is om alleen studies op te nemen met een (quasi-) experimentele opzet is bij sommige thema's gebleken dat er interessante ontwikkelingen hebben plaatsgevonden die hoewel zij technisch niet tot de doelgroep behoren wel zijn opgenomen in dit overzicht.

In de rest van het rapport worden de verschillende thema's per hoofdstuk behandeld. Het rapport zal worden afgesloten met enkele algemene conclusies en opmerkingen.

² In de bijlage zijn als referentie enkele overzichtstabellen uit deze publicaties opgenomen.

2 Onderwijspersoneel

Het thema onderwijspersoneel is, zoals in de inleiding is aangegeven, opgedeeld in vijf onderdelen: 1) de relatie tussen de kwaliteit van leraren en de kwaliteit van het onderwijs, 2) de relatie tussen de kwaliteit van leraren en hun beloning, 3) de relatie tussen de kwaliteit van het onderwijs en de beloning van onderwijspersoneel, 4) de mogelijkheden tot verbetering van de kwaliteit van leraren en 5) de rol van ondersteunend onderwijspersoneel.

De onderlinge samenhang van deze lijst is duidelijk maar niet compleet. De effecten van een verandering van de beloning voor onderwijzend personeel zijn niet los te zien van de institutionele omgeving. Dit is uiteraard voor elk onderzoek belangrijk alleen spelen er hier enkele zeer belangrijke factoren een rol die makkelijk worden vergeten. Voor het implementeren van financiële prikkels (in de vorm van hogere beloning) is het noodzakelijk dat er een motivatie is voor de directie (het schoolbestuur) om de instelling te positioneren en te onderscheiden van overige onderwijsinstellingen. Deze motivatie is echter sterk afhankelijk van de vraag vanuit de leerlingen en nog belangrijker in het geval van primair- en voortgezet onderwijs, vanuit de ouders. Alleen wanneer de kwaliteit van een school daarin een (mede-)bepalende factor wordt en de 'vraag' naar onderwijs (bijvoorbeeld gemeten naar het aantal aanmeldingen) hiervan afhangt wordt deze keten in gang gezet. Daarmee is niet gezegd dat scholen niet een intrinsieke drang hebben om de kwaliteit van het onderwijs te verbeteren. Het punt is dat dergelijke op marktwerking georiënteerde interventies alleen werken wanneer de consumenten, in dit geval de leerlingen en hun verzorgers, ook daadwerkelijk hun 'marktkracht' aanwenden. De effectiviteit van prikkels moet in dit licht worden beschouwd: ouders moeten gevoelig zijn voor kwaliteitsverschillen, de scholen moeten op hun beurt weer gevoelig zijn voor deze signalen welke omgezet worden in een hogere kwaliteit bijvoorbeeld door financiële (of niet-monetaire) prikkels voor leerkrachten die daar op hun beurt ook gevoelig voor moeten zijn. Alvorens de verschillende subonderdelen te behandelen worden eerst drie studies behandeld die deze materie als onderwerp hebben. Hoewel de eerste en de derde geen (quasi-) experimentele studies zijn worden deze toch behandeld om een indicatie te geven waar verder onderzoek op gericht zou moeten zijn.

Concurrentie in onderwijs

Beleidsmakers in de UK pleiten voor meer keuzevrijheid van ouders met betrekking tot de schoolkeuze van hun kind(eren). Men denkt dat meer keuzevrijheid leidt tot meer concurrentie tussen scholen en dat als gevolg daarvan de kwaliteit van het onderwijs omhoog gaat. In Gibbons *et al* (2006) wordt onderzocht of dit empirisch wordt ondersteund. Aan de hand van gegevens over de basisschoolpopulatie in Zuidoost Engeland worden twee indices gedefinieerd: een index voor de mate van keuzevrijheid voor ouders (*keuze-index*) en een index voor de mate van concurrentie tussen scholen (*concurrentie-index*). De *keuze-index* geeft het aantal scholen weer waartoe een leerling toegang heeft gebaseerd op de afstand tussen het woonadres en het schooladres. De *concurrentie-index* geeft voor elke school het gemiddeld aantal scholen weer waartoe studenten toegang hebben. Vervolgens wordt een empirisch model geschat waarin de uitkomsten van een toets die 10/11-jarigen maken worden verklaard door de toetsresultaten die

dezelfde studenten op 6/7-jarige leeftijd hebben gemaakt, de *keuze*-index, *concurrentie*-index en andere controle variabelen. Deze variabelen zijn niet exogeen en er is dus ook geen sprake van een (quasi-) experimentele opzet.

Uit de analyse blijkt dat leerlingen die kunnen kiezen uit meerdere scholen het niet beter doen dan leerlingen die een beperktere keuze hebben. De algemene conclusie is dat op basis van deze analyse niet gesteld kan worden dat meer keuzevrijheid en meer concurrentie leidt tot een verbetering van de prestaties van leerlingen.

In Lavy (2006) wordt het effect van een vrijere schoolkeuze op de prestaties van leerlingen aan het eind van de middelbare-schoolcarrière (drop-out, slagingspercentages eindexamen, gemiddelde cijfers, etc.) onderzocht. Hierbij wordt een onderscheid gemaakt tussen het individuele effect op de leerling (wordt hij/zij beter van de keuze voor een ander district?) en het algemene evenwichtseffect op de districten (is er een betere mix van leerlingen die tot gemiddeld betere resultaten leidt?). Lavy maakt in het onderzoek gebruik van een verandering in het onderwijsbeleid in Tel Aviv. Tot 1994 kende de stad een systeem waarin elke leerling op de middelbare school verplicht naar een bepaalde school werd gestuurd; tweederde in het eigen district en eenderde in een ander district. In 1994 werd dit systeem in één district vervangen door een met iets meer vrije keuze. Men mocht nu zelf kiezen uit vijf van tevoren aangewezen scholen, drie in het eigen district en twee in een ander district.

Het algemene evenwichtseffect op de districten wordt bepaald m.b.v. een controlegroep uit een direct naastgelegen district waarin de wetswijziging nog niet was ingevoerd, aan de hand van panelregressie en 'difference-in-differences'. De individuele effecten worden bepaald door middel van IV-regressie, m.b.v. instrumentele variabelen gebaseerd op verschillen in openbaar vervoermogelijkheden en het hebben van broers/zusjes op scholen buiten het eigen district.

De hypothese die door Lavy wordt getest is dat de schoolresultaten door de vrijere keuze beter worden. Dit komt enerzijds omdat scholen gedwongen worden onderlinge competitie te voeren voor leerlingen en anderzijds omdat een betere 'matching' mogelijk wordt tussen leerlingen en scholen. Beide effecten kunnen zowel op individueel als op districtsniveau optreden. Uiteindelijk vindt Lavy inderdaad het algemeen evenwichtseffect van vrijere keuze: het district haalt in allerlei opzichten betere prestaties na invoering van de vrijere keuze. Een individueel effect van een schoolgang buiten het eigen district wordt niet gevonden.³

De Wolf en Janssens (2005) geven een overzicht van de studies die zijn verricht naar de gevolgen van controlemechanismen in het onderwijs in verschillende landen. Aan de hand van de

³ Voor verdere studies naar de effecten van keuze en concurrentie, zie Hoxby (2000) (behandeld in Webbing, 2005) en de reactie van Rothstein (American Economic Review, forthcoming) en het weerwoord van Hoxby hierop. Hoxby vindt enorm grote effecten van meer keuze op prestaties (positief) en kosten (negatief). Keuze wordt gemeten als een index op basis van aantal en concentratie van districten binnen een stedelijk gebied. Variatie in deze index wordt geïnstrumenteerd met het aantal grote en kleine rivieren dat door het gebied stroomt. Rothstein laat zien dat de uitkomsten uiterst gevoelig zijn voor de manier waarop het aantal rivieren precies wordt gemeten. Ook de studie van Cullen, Jacobs en Levitt uit 2006 gaat hierover. In deze studie (hun werk wordt behandeld in Van der Steeg (2006) en Webbing (2005) wordt gebruik gemaakt van variatie in keuze die tot stand komt doordat bij overaanmelding op middelbare scholen in Chicago erom wordt geloot wie wordt toegelaten. De winnaars van deze loterijen hebben een grotere keuzeset dan de verliezers, toch zijn hun prestaties niet beter.

behandelde onderzoeken komen zij onder meer tot de conclusie dat er geen duidelijk beeld is over de causale effecten van schoolinspecties of prestatie-indicatoren op de kwaliteit van scholen/onderwijs. Opmerkelijk genoeg komt hier ook naar voren dat ouders, wanneer zij een school voor hun kinderen kiezen, vrijwel geen gebruik maken van openbare kwaliteitsindicatoren. Tegelijkertijd komt uit de literatuur naar voren dat deze indicatoren door bestuurders en onderwijzend personeel als belangrijk worden bestempeld. Ook blijkt dat controle leidt tot strategisch gedrag van scholen, waarbij men moet denken aan tactieken als *window-dressing*, *teaching to the test*, frauderen en het aanpassen van de groep leerlingen bij wie een toets wordt afgenomen om zo de uitkomsten positief te beïnvloeden. Over het algemeen, zo geven de auteurs aan, wordt te weinig rekening gehouden met de causaliteit en wordt er weinig met controlegroepen gewerkt. Aan het eind van het artikel vermelden zij enkele suggesties om dit te ondervangen, waaronder een opzet voor willekeurige selectie van onderwijsinspecties.

In de rest van dit hoofdstuk worden de vijf subonderdelen van het thema ‘Onderwijspersoneel’ achtereenvolgens behandeld.

De relatie van de kwaliteit van leraren en de kwaliteit van het onderwijs

De kwaliteit van leraren is een zeer belangrijke determinant van de prestaties van leerlingen. Een studie verricht in de Amerikaanse staat Texas laat zien dat er grote verschillen in de bijdragen van docenten aan deze prestaties. Uit de in Cornet *et al* (2006) aangehaalde studies blijken twee factoren succesvolle instrumenten te kunnen zijn om de prestaties van leraren te verbeteren: scholing van leraren en teambeloning. De algemene boodschap is echter dat hier verder onderzoek noodzakelijk is.

Het verhogen van de kwaliteit van leraren (om de prestaties van leerlingen te verbeteren) wordt in Cornet *et al* (2006) als een zeer kansrijke beleidsoptie getypeerd. Hierbij speelt de vraag in hoeverre de uitkomsten van een interventie bijdragen aan de het verbeteren van het maatschappelijke rendement een grote rol. Het huidige institutionele systeem bevat weinig prikkels ter verhoging van de kwaliteit van leraren. De beloning van leraren is vooral afhankelijk van de ervaring en in veel mindere mate van prestaties. De uitkomsten van de in Cornet *et al* (2006) aangehaalde onderzoeken laten zien dat scholing en teambeloning kosteneffectiever zijn dan bijvoorbeeld klassenverkleining. Tevens geven zij aan dat de kennis over de precieze determinanten nog beperkt is waardoor de mogelijkheid bestaat dat er betere alternatieven zijn die de kwaliteit van leraren verbeteren.

Clotfelter *et al* (2007) onderzoeken wat de invloed is van de kwaliteit van leraren op de prestaties van leerlingen in het primair onderwijs. Door middel van het koppelen van verschillende datasets hebben de auteurs een dataset gemaakt (panel data) waarin gegevens over testcores van de leerlingen (zowel rekenen als taal), de kwalificaties en achtergrondfactoren van leraren, de karakteristieken van de klas en de achtergrondfactoren van leerlingen. Ondermeer met fixed-effect regressies worden de verschillende kenmerken geschat op de testcores van leerlingen. Het voordeel hiervan is dat er kan worden gecorrigeerd voor de niet-waarneembare kenmerken van leerlingen. Dat is nodig, omdat leraren niet random over klassen zijn verspreid. De kenmerken van leerlingen (en dus zeer waarschijnlijk ook de niet-waarneembare kenmerken) spelen hierbij een rol. Kort door de bocht gezegd werken de beste leraren in de beste klassen. In de fixed-effect regressies kan hiervoor worden gecorrigeerd.

Uit het onderzoek komt onder meer naar voren dat de achtergrondkenmerken van leerlingen (zoals het opleidingsniveau van de ouders) meer invloed hebben op de taalscores dan op rekenscores. Ten aanzien van de leraren blijkt dat hun kwalificaties meer invloed op rekenen hebben dan op taal. Hoe meer ervaring een leraar heeft, des te beter de testcores van de leerlingen. Leraren die als zij-instromers zijn begonnen hebben ten opzichte van leraren met een normaal diploma een negatief effect op de testcores. Deze uitkomsten impliceren een aantal mogelijke beleidsinitiatieven zoals het aantrekkelijker maken van langer doorwerken in het onderwijs; oppassen met zij-instromers en proberen goede leraren ook voor minder goede klassen te krijgen. In Clotfelter *et al* (2006) is deze laatste suggestie onderzocht en zal hieronder worden uitgewerkt.

Wanneer het gaat om de effecten van de kwaliteit van leraren, dan is er thans geen Nederlands onderzoek voorhanden. Om deze effecten te bepalen zal men echter toch met Nederlandse data aan de slag moeten. De status is dat het een ‘witte vlek’ is maar rijp voor kansrijke experimenten.

De relatie van de kwaliteit van leraren en hun beloning

In Clotfelter *et al* (2006) wordt de relatie tussen de kwaliteit van leraren en hun compensatie behandeld. De auteurs menen dat de aanzienlijke aandacht voor deze relatie voortkomt uit het verlangen om het aantal kwalitatief goede individuen dat een carrière als leraar wil opnemen te vergroten en om te voorkomen dat door verloop en uitval de meest gekwalificeerde individuen geen leraar meer blijven. Een andere motivatie, die de kern vormt van hun onderzoek, is de zorg over de ongelijke distributie van effectieve leraren tussen scholen. Deze groep heeft de neiging om zich richting de scholen te bewegen alwaar de kinderen uit betere achtergrondmilieus komen, zelfs wanneer de salarissen daar niet hoger zijn.

Het paper kijkt naar een programma dat van 2001/02 tot 2003/04 actief was in de staat North Carolina waarbij jaarlijkse bonussen (tot 1.800 dollar) werden uitgekeerd aan gecertificeerde leraren (wiskunde, natuurkunde/scheikunde en bijzonder onderwijs) die werkten in de ‘mindere’ scholen (dwz met veel achterstandskinderen). Het doel van het programma was om het voor deze scholen makkelijker te maken om gekwalificeerde leraren aan te trekken en te behouden. Om het succes van het programma na te gaan wordt gekeken naar het verloop van de leraren per school. De controlegroep bestaat uit scholen die net buiten de criteria van het programma vielen. De resultaten geven aan dat de omvang van de bonus groot genoeg was om het verloop te laten afnemen. Tevens bleek dat het meer effectief was voor leraren met tien tot dertig jaar werkervaring en een grotere impact had op wiskundeleraren.

Kan de kwaliteit van het onderwijs worden verbeterd door een andere manier van beoordelen van leraren en de daaraan verbonden beloning (in plaats van de traditionele beloning op basis van opleiding en ervaring)? Deze vraag staat centraal in het artikel van Jacob & Lefgren (2006) waarbij de kwaliteit van leraren (in groep 2 t/m 8 van basisscholen in de Verenigde Staten) aan de hand van drie indicatoren wordt uitgevoerd:

- De toegevoegde waarde van leraren, gemeten op basis van testcores van leerlingen
- Beoordeling van functioneren van leraren door het schoolhoofd

- Tevredenheid van ouders, uitgedrukt in het aantal malen dat ouders een voorkeur uitspreken voor een leraar bij de toewijzing van een leerling

Deze drie indicatoren worden op verschillende manieren vastgesteld. De toegevoegde waarde wordt gemeten door middel van een econometrisch model voor de verklaring van testcores, waarin (de toegevoegde waarde van) de leraar één van de verklarende factoren is, naast allerlei kenmerken van de leerling, de school en het tijdstip waarop is gemeten. Dit model is geschat met de gewone kleinste kwadraten methode. De beoordeling van leraren door het schoolhoofd is gemeten via een enquête onder schoolhoofden, waar elke leraar is beoordeeld op een tiental aspecten, plus een algemene beoordeling. De tevredenheid van ouders wordt gemeten op basis van gegevens van scholen.

In het artikel van Jacob & Lefgren (2006) wordt nergens ingegaan op causaliteit, alleen op samenhang. Maar op basis van die samenhang wordt wel geprobeerd de vraag te beantwoorden of een andere manier van belonen (op basis van het oordeel van het schoolhoofd in plaats van op basis van opleiding en ervaring) ertoe zou kunnen leiden dat leraren beter presteren. Impliciet wordt dus verondersteld dat een financiële prikkel leidt tot beter presteren (causaliteit 1) en dat een betere prestatie leidt tot een betere beoordeling door het schoolhoofd (causaliteit 2), hetgeen nauwkeuriger zou zijn dan de manier waarop opleiding en ervaring leiden tot beter presteren (causaliteit 3).

Toekomstige prestaties van leerlingen (testcores) kunnen beter worden voorspeld aan de hand van de kwaliteit van leraren zoals beoordeeld door schoolhoofden, dan zoals zou blijken uit opleiding en ervaring of simpelweg het loon van leraren. Hoewel de toegevoegde waarde van docenten, zoals vastgesteld op basis van testcores in het verleden, nog beter in staat is om de toekomstige prestaties van leerlingen goed te voorspellen, zijn schoolhoofden net zo goed in staat om de beste en slechtste leraren te identificeren. Als het gaat om de tevredenheid van ouders, dan is het oordeel van schoolhoofden over leraren zelfs een betere inschatting dan de toegevoegde waarde van leraren op basis van testcores. Kennelijk vinden ouders ook andere aspecten van leraren belangrijk dan alleen het kunnen verbeteren van testcores en kunnen schoolhoofden die andere aspecten goed inschatten.

Deze resultaten impliceren dat de kwaliteit van het onderwijs kan worden verbeterd via de kwaliteit van de leraren, wanneer het oordeel van schoolhoofden een grotere rol krijgt in de beloning van leraren. Daarbij moet wel rekening gehouden worden met het feit dat schoolhoofden vooral goede en slechte leraren kunnen onderscheiden, niet noodzakelijk leraren die marginaal verschillen in kwaliteit. Ook bestaat het gevaar van fouten in de beoordeling. Zo blijkt dat vooral de kwaliteit van mannelijke leraren en leraren met een tijdelijk contract te laag werd ingeschat door schoolhoofden.

De relatie van de kwaliteit van het onderwijs en de beloning van onderwijspersoneel

Het onderzoek van Lavy (2005) presenteert de effecten van een individuele monetaire premie gebaseerd op een experimenteel rankorder toernooi onder leraren Engels en wiskunde in Israël. Leraren konden een cash bonus verdienen door de examenprestaties van hun leerlingen te verbeteren. De uitkomsten van dit experiment wijzen op significant positieve verbeteringen van de onderwijsprestaties van de leerlingen waarbij niet alleen de eindresultaten verbeterden maar

ook het aantal leerlingen dat voor een examen opging toenam. Uit een analyse van de aanpak van leraren blijkt dat deze verbeteringen het gevolg waren van veranderingen in onderwijsmethoden van leraren evenals een toename van naschools onderwijs en meer ontvankelijkheid voor de behoeftes van leerlingen.

Figlio en Kenny (2006) trachten als eerste een systematische documentatie te geven over de relatie tussen individuele prestatie afhankelijke beloning voor leraren en de leerling-prestaties in de Verenigde Staten. De auteurs vinden een positieve associatie tussen de individuele prestatiebeloning en de onderwijsprestaties van leerlingen. Zij identificeren twee interpretaties van deze uitkomst. Allereerst is het mogelijk dat de financiële prikkel de leraren ertoe heeft geleid om hun inzet te verhogen. Als tweede interpretatie vermelden zij dat het ook mogelijk is dat dergelijke beloningsprogramma's juist op scholen worden geïmplementeerd die effectiever zijn op een manier die niet makkelijk meetbaar en niet d.m.v. de data makkelijk te identificeren is. De auteurs geven aan dat experimenteel onderzoek deze zaak zal moeten beslechten.

In 1997 stapte de stad Chicago over op een systeem waarbij de openbare scholen voortaan worden afgerekend op basis van de onderwijsprestaties van hun leerlingen. De achterliggende gedachte is dat een dergelijk systeem zowel leerlingen (door mogelijke sancties als *summer school* of doubleren) als leraren en schoolbesturen een prikkel geven om de onderwijsprestaties te verbeteren. Daarnaast helpt het bij het identificeren van mogelijke probleemscholen. Jacob (2005) gebruikt de introductie van dit beleid, vaak aangeduid als *high-stakes testing*, om na te gaan of dit beleid de onderwijsprestaties van leerlingen ook daadwerkelijk verbetert. Daarnaast onderzoekt hij ook welke factoren hierbij een rol hebben gespeeld en of leraren en schoolbestuurders sinds de introductie strategisch gedrag zijn gaan vertonen. Dit laatste is mede ingegeven door de provisie dat indien een school consequent lage scores behaalt er ingegrepen kan worden in de vorm van ontslag of herplaatsing van leerlingen en bestuurders.

De uitkomsten van het onderzoek wijzen op een verbetering van de prestaties van leerlingen op het gebied van wiskunde en leesvaardigheid. Onder de jongere leerlingen leidde het beleid vooral tot een toename van specifieke 'toetskennis'. Dit laatste wil zeggen dat leerlingen vooral onderwijs krijgen dat gericht is op het behalen van een goede score op de toetsen waarbij andere niet getoetste aspecten van het curriculum minder aandacht kregen. Een andere factor die het positieve resultaat beïnvloedde was dat leerlingen meer hun best zijn gaan doen. Ook blijkt uit het onderzoek dat er sprake is van strategisch gedrag van leraren. Dit uitte zich onder meer in een toename van het aantal plaatsen in de leergang voor bijzonder onderwijs, het preventief laten doubleren. De algemene conclusie van Jacob is dat *high-stakes testing* in potentie een grote bijdrage kan leveren aan de onderwijsprestaties maar dat het instrument wel met enige zorg zal moeten worden benaderd.

Uit het gepresenteerde overzicht valt op te maken dat onderzoek naar de effectiviteit van (financiële) prikkels voor leraren vaak ook een analyse van mogelijke fraude bevatten. (zie Lavy (2005) en Jacob (2005)). Hieronder wordt een onderzoek weergegeven naar de wijze waarop eventuele fraude kan worden opgespoord.

In Jacob en Levitt (2003) beschrijven de auteurs een experimenteel onderzoek naar frauduleus gedrag van leerkrachten (en schooladministraties) bij centrale vaardigheidstoetsen voor leerlingen

in het primair en het secundair onderwijs. De auteurs hadden in een eerder onderzoek algoritmes ontwikkeld die fraude kunnen opsporen in antwoordenreeksen van leerlingen. Die algoritmes gaan op zoek naar bijvoorbeeld ongewone aantallen van identieke antwoordreeksen van leerlingen in één klas of binnen de antwoorden van één leerling naar opmerkelijke aantallen goede antwoorden op moeilijke vragen in combinatie met foute antwoorden op eenvoudige vragen. De effectiviteit van de algoritmes tonen de auteurs aan door te kijken naar prestaties van de leerlingen een jaar later. Het blijkt dat de winst in leerprestaties in verdachte klassen met grote kans verdampt in het jaar daarna. Anders gezegd: klassen waarin relatief veel leerlingen een onverwacht hoge uitslag laten volgen door een onverwachte lage uitslag het jaar daarna, zijn de klassen waarin veel verdachte antwoordenreeksen worden waargenomen.

Aan de hand van deze algoritmes selecteerden de auteurs een aantal verdachte klassen voor de hertoets, die jaarlijks in Chicago steekproefsgewijs wordt uitgezet. De hertoets kreeg vorm als een experiment met een controlegroep van willekeurige klassen. Uiteraard konden de leerkrachten bij de hertoets niet frauderen. De uitkomsten zijn overtuigend. In de controlegroep wordt nauwelijks verschil waargenomen tussen de aanvankelijke toets en de hertoets. In de verdachte klassen zijn de hertoetsuitslagen veel lager en daaruit wordt geconcludeerd dat in 3 tot 5 procent van alle klassen van openbare scholen in Chicago wordt gefraudeerd met de centrale vaardigheidstoets.

Tot besluit van het artikel gaan de auteurs op zoek naar welke van de verdachte antwoordenreeksen het beste een fraudeklas aanduiden. Verreweg de beste indicator is die waarbij wordt gekeken naar leerlingen met veel moeilijke vragen goed en makkelijke vragen fout. Ook een hoge correlatie tussen de antwoordreeksen van leerlingen binnen een klas is een sterke aanwijzing. Het vaak voorkomen van identieke blokken binnen de antwoordenreeksen van leerlingen blijkt geen sterke aanwijzing voor fraude.

De mogelijkheden van het verbeteren van de kwaliteit van leraren

Het blijkt in de onderzoekspraktijk moeilijk om de kwalitatief goede leraren te identificeren. Dit is een lastig probleem omdat dergelijke leraren veel meer geneigd zijn om mee te doen aan additionele trainingen dan minder kwalitatieve leraren. Hierdoor zou het beeld kunnen ontstaan dat additionele training voor leraren leidt tot betere schoolprestaties van leerlingen. Webbink (2005) behandelt twee studies waarin de kwaliteit van leraren wel wordt geïdentificeerd. De uitkomsten geven geen gewicht aan het idee dat extra geld voor leraren altijd leidt tot betere prestaties. De belangrijkste vinding van de ene studie is dat investeringen in de ontwikkeling van leraren geen verbetering van de prestaties van leerlingen op basisscholen met veel kinderen uit arme gezinnen hebben gegenereerd. De beperkte omvang van de investeringen wordt aangewezen als belangrijkste verklaring hiervoor. De tweede studie heeft geen significante effecten gevonden.

De rol van ondersteunend onderwijspersoneel

Op een school werken meer mensen dan alleen onderwijzers. In het beroemde experiment STAR naar de effecten van klassenverkleining in Tennessee was één van de *treatments* de toevoeging van

een onderwijsassistent.⁴ Uit deze analyse bleek dat leerlingen in klassen met onderwijsassistent het niet beter of slechter doen dan klassen zonder zo'n assistent. Maar afgezien van onderwijsassistenten is het effect van (niet onderwijzend) onderwijspersoneel nog nooit degelijk onderzocht. Dit mag dan ook met recht als 'witte vlek' worden bestempeld in de kennis over de impact van onderwijspersoneel op de maatschappelijke baten van onderwijs.

⁴ Krueger, A.B., 1999, Experimental estimates of education production functions, *Quarterly Journal of Economics*, 114 (2), p. 497-532. (zie voor een behandeling van deze study o.a. Webbink, 2005)

3 Voortijdig schoolverlaters / leer-werkplicht

Voortijdig schoolverlaten wordt als één van de grootste problemen van het Nederlandse onderwijs beschouwd. Zoals Van der Steeg en Webbink (2006) aangeven in hun overzichtspublicatie “Voortijdig schoolverlaten in Nederland: omvang, beleid en resultaten” is niet bekend of het Nederlandse beleid bijdraagt aan een vermindering van voortijdig schoolverlaten. De auteurs gebruiken de uitkomsten van buitenlandse (quasi-)experimentele studies om een beeld te schetsen van mogelijke succesvolle interventies.⁵ De effectiviteit van programma’s ter bestrijding van uitval, zo blijkt uit de behandelde studies, geeft een gemengd beeld. Veel van de onderzochte programma’s hebben geen of weinig effect op het verminderen van voortijdig schoolverlaten. Wanneer er wel sprake was van positieve effecten op uitvalgedrag werden deze hoofdzakelijk gevonden bij programma’s die financiële prikkels voor scholen, leraren of leerlingen hadden ingebouwd. Daarnaast blijkt uit deze literatuur dat intensieve projecten (zoals het toewijzen van individuele mentoren aan risicoleerlingen) positieve resultaten kunnen genereren. Het algemene beeld dat uit de internationale literatuur naar voren komt is dat een preventieve aanpak effectiever is dan een curatieve aanpak.

Het probleem voor de Nederlandse situatie is dat veel van de evaluatiestudies geen gebruik hebben gemaakt van controlegroepen. Deze groepen kunnen bestaan uit een groep leerlingen (of scholen) die geen middelen (verbonden met voortijdig schoolverlaten) hebben ontvangen of juist extra middelen hebben gekregen. Door het ontbreken van een controlegroep is nu niet bekend of de situatie beter of slechter is dan wanneer de betreffende instrumenten niet zouden zijn ingezet. Met andere woorden, een ‘counterfactual’ ontbreekt. De auteurs geven tegelijkertijd aan dat het Nederlandse beleid wel mogelijkheden biedt om deze witte vlek weg te werken: “Een inventarisatie van het bestaande vsv-instrumentarium in Nederland levert twee instrumenten op die mogelijksterwijs in aanmerking komen voor een evaluatie op basis van een natuurlijk experiment: de ESF/vsv-subsidies en de trajectbegeleidingscomponent van de RMC-functie.” (Van der Steeg & Webbink, 2006, p. 68)

Sinds de publicatie van Van der Steeg & Webbink (2006) zijn geen nieuwe onderzoeken op dit thema verschenen.

⁵ Zie tabel A in de bijlage voor een overzicht van deze studies en de belangrijkste uitkomsten

4 Voor- en vroegschoolse educatie

Wanneer het gaat om onderwijs dan is de boodschap die uit alle onderzoek naar voren komt: hoe eerder hoe beter. Vrijwel elke evaluatie vindt sterke positieve effecten, ook op latere leeftijd, van een vroege aanvang.⁶

Leuven *et al* (2006) onderzoeken het effect van het uitbreiden van de toegang tot vroegschoolse onderwijs op de onderwijsprestaties in het begin van de schoolcarrière. Het onderzoek is quasi-experimenteel van opzet, gebaseerd op twee eigenschappen van het Nederlandse onderwijssysteem. Enerzijds maken zij gebruik van het feit dat er een continue inschrijving mogelijk is. Kinderen mogen direct na hun vierde verjaardag in het onderwijs participeren en hoeven dus niet te wachten tot de aanvang van het eerstvolgende schooljaar. Daarnaast geldt dat kinderen die hun verjaardag vieren vlak voor, tijdens of na de zomervakantie in dezelfde klas worden gezet. Deze twee eigenschappen creëren voldoende exogene variatie op de maximale duur van onderwijs om de effecten daarvan te identificeren op de uitkomsten van twee cognitieve (Cito) toetsen. De analyse vindt significant positieve effecten voor kinderen uit achterstandsmilieus. Bij een verlenging van één maand verbeteren de testcores voor zowel taal als rekenen waarbij het effect van de eerste iets groter is dan de tweede.

Veel onderzoek is inmiddels verricht op basis van het Amerikaanse 'Head Start' programma. Dit programma geeft onder meer voorschools onderwijs aan kinderen uit achterstandsgezinnen. De effecten van dit programma zijn al meerdere malen bestudeerd. Het nieuwste onderzoek van Currie en Neidell (2007) behandelt de claims dat het programma teveel geld besteedt dat niet direct gerelateerd is aan kinderen. Door gebruik te maken van verschillende databases, sommigen reiken terug tot de jaren zeventig, kan de loopbaan van kinderen die aan het programma hebben deelgenomen over tijd worden geanalyseerd. Uit het onderzoek komt naar voren dat deze kinderen hogere cijfers op examens/ toetsen haalden wanneer de uitgaven aan het programma hoger waren. Wanneer de per capita uitgaven constant worden gehouden blijkt dat kinderen in programma's waar meer van het budget opging aan kindgerelateerde uitgaven, minder gedragsstoornissen hebben en minder vaak blijven zitten.

⁶ Zie o.a. tabel C in de bijlage

5 Schaalgrootte

Zoals in de inleiding reeds is aangegeven is er vrijwel geen onderzoek uitgevoerd naar de impact van de grootte van de onderwijsinstelling (vestigingen) op onderwijsprestaties waarbij gebruik is gemaakt van een (quasi) experimentele opzet. Dat wil echter niet zeggen dat er in het geheel geen onderzoek naar is gedaan. In 2002 verscheen het proefschrift “Schaal in het primair onderwijs” van Van de Venne. Recenter heeft de Onderwijsraad (2005) een advies uitgegeven over deze materie met als titel “Variëteit in schaal, keuzevrijheid, sociale samenhang en draagvlak bij grote onderwijsorganisaties”.

Het advies van de Onderwijsraad begint met de observatie dat het overheidsbeleid in de onderwijssector de afgelopen jaren sterk geënt was op fusies van instellingen en besturen. Dit heeft onder meer geleid tot een toename in de schaalgrootte van instellingen. Deze schaalvergroting, zo was het idee, zou de doorstroom van leerlingen en de kwaliteit van het onderwijs ten goede komen. Tegelijkertijd heeft dit beleid, zo merkt de Onderwijsraad op, ervoor gezorgd dat er tevens ongewenste effecten zijn voor de keuzevrijheid van ouders en leerlingen en de sociale cohesie binnen opleidingen. Deze gevolgen vormen de aanleiding van het advies. In de onderstaande tabel zijn de verschillende mogelijke voor- en nadelen van ‘aanzienlijke’ schaalgrootte van onderwijsinstellingen weergegeven.

Tabel 5-1 Mogelijke voor- en nadelen van aanzienlijke schaalgrootte

Voordelen	nadelen
<ul style="list-style-type: none"> - breed onderwijsaanbod; meer mogelijkheden voor differentiatie en specialisatie; - doelmatigheidswinst (in eerste aanleg); - scholengemeenschap met verschillende schooltypen kan integratie bevorderen; - groter beleidsvoerend vermogen; professionaliteit/expertisebundeling; - meer financiële armslag; risicospreiding en grotere investeringscapaciteit; - als speler op hoger niveau mee kunnen doen (geldt voor alle sectoren van basisscholen tot hogescholen); en - meer loopbaanperspectieven en specialiseringsmogelijkheden voor het personeel. 	<ul style="list-style-type: none"> - doelmatigheidsverlies bij doorgesloten schaalvergroting (het mechanisme van budgetmaximalisatie); - beperking van keuzemogelijkheden qua aanbod voor deelnemers/ouders; - risico van anonimiteit en verminderde betrokkenheid van ouders, deelnemers, personeel en afnemers; - verminderde invloed op de besluitvorming van ouders, deelnemers en personeel; - grote concentratie van leerlingen (met name in het voortgezet onderwijs en het beroepsonderwijs) kan overlast voor omgeving veroorzaken; en - ontbreken van legitimatie en controle in het geval van lokale of regionale monopolies.

Bron: Onderwijsraad (2005, p. 12)

In het advies concludeert de Onderwijsraad dat ‘fusies tussen onderwijsinstellingen de monopolievorming in een bepaald gebied kunnen versterken en daarmee de keuzevrijheid voor de deelnemer belemmeren. Met name in de bve-sector (beroepsonderwijs en volwasseneneducatie) en het hbo (hoger beroepsonderwijs) zijn er regio’s met nog maar een of twee aanbieders.’ (Onderwijsraad, 2005, p.9) Ook blijkt dat ouders, deelnemers en personeel een grotere afstand bespeuren tussen hen en de besturen van de grote onderwijsinstellingen. Ook wordt vermeld dat afnemers (gedefinieerd als aanbieders van vervolgoopleidingen en toekomstige werkgevers) minder binding hebben met het onderwijsaanbod waardoor de aansluiting minder goed verloopt. Het uiteindelijke advies aan de minister van Onderwijs is om een fusietoets in te

stellen zoals deze reeds in andere sectoren wordt toegepast door de Nederlandse Mededingingsautoriteit. Hierbij dient de aanwezige keuzevrijheid voor ouders en deelnemers centraal te staan. Daarnaast geeft de Onderwijsraad het advies om de kring van belanghebbenden bij scholen- en opleidingsbeslissingen uit te breiden om de legitimatie alsmede de degelijkheid van het onderwijsbestuur te waarborgen.

In haar proefschrift behandelt Van de Venne (2002) de invloed van schaalverschillen en schaalverandering op de organisatie-effectiviteit in het basisonderwijs. Daarnaast bekijkt zij ook hoe een eventuele beïnvloeding in de praktijk tot uiting komt. Het onderzoek maakt gebruik van referentiesteekproef van het PRIMA-cohort 1996 met additionele scholen uit de PRIMA-steekproef en additionele survey-data. Onder de uitkomsten van het onderzoek is onder meer de positieve relatie tussen schaalgrootte en standaardisatie van leerprocessen, het gebruik van leerlingvolgsystemen. De effectiviteit van schoolorganisatie wordt echter maar voor een klein deel bepaald door schaalvergroting. De belangrijkste factor, zo blijkt uit het onderzoek, is de sturing door de schoolleider.

6 Examinering

Uit de wetenschappelijke literatuur komt in toenemende mate het beeld naar voren dat de aanwezigheid van centraal gehouden eindexamens (CEE) sterk positief gerelateerd is met de onderwijsvoortgang van leerlingen en studenten. Uit het overzicht van Bishop (2002) komt duidelijk naar voren dat de aanwezigheid van gestandaardiseerde centrale eindexamens samengaat met hogere onderwijsprestaties. Dit blijkt onder meer uit een analyse tussen landen met een vergelijkbare economische ontwikkeling die dit systeem niet kennen en landen die wel een dergelijke centrale examinering hanteren. De vraag is nu of deze bevinding ook overeind blijft binnen een evaluatie waarbij de endogeniteit is geadresseerd.

Jürges *et al* (2005) onderzoeken de invloed van een externe centrale examinering waarbij zij gebruik maken van de institutionele eigenschappen van het federale onderwijssysteem in Duitsland als bron van exogene variatie. Deze variatie creëert de mogelijkheid om het effect van de introductie van een CEE op de schoolprestaties te identificeren. In hun onderzoek concentreren zij zich op de examens aan het einde van het 'lage' voortgezet onderwijs (*Hauptschule* en *Realschule*). In de meeste *länder* krijgen leerlingen na afronding van dit niveau een certificaat. Enkele kennen echter een centraal afsluitend examen in drie tot vier vakken. Doordat de leerlingen zelf aangeven in welke vakken zij het examen willen afleggen is er een mogelijk zelfselectie effect. De onderzoekers beperken zich in hun analyse op wiskunde wat samen met Duits de twee verplichte vakken zijn. Bij het bepalen van het effect van een centrale examinering maken zij gebruik van de TIMSS data.⁷ De gegevens voor de overige vakken worden aangewend om een additionele *difference-in-differences* per vak te schatten. Door optimale benutting van de verschillende gegevens fungeert elke leerling als zijn/haar eigen controlegroep.⁸ Om optimaal gebruik te maken van de exogene variatie kijken zij niet alleen naar verschillen tussen de deelstaten met en zonder centrale examinering maar ook binnen enkele staten waar beide systemen naast elkaar bestaan en voor-en-na vergelijking. Bij deze laatste strategie worden de leerlingen van deelstaten bekeken voor en na de invoering van een centraal examen. Het gemiddelde effect van de aanwezigheid van een centraal examen op de scores voor wiskunde is een verbetering van 0,12 standaarddeviatie. De algemene conclusie is dat de introductie van een centrale examinering de gemiddelde schoolprestaties zal laten stijgen. Verder merken de auteurs op dat deze oplossing vergeleken met veel andere interventies zoals het verkleinen van klassen aanzienlijk kosteneffectiever is.

Een interessante vraag is hoe de effecten van centrale examinering zijn verdeeld over de leerlingen. Wößmann (2004) tracht de heterogeniteit van het effect van externe eindexamens op de onderwijsvoortgang van leerlingen te schatten. Hierbij wordt een onderscheid aangebracht tussen drie dimensies: bekwaamheid van leerlingen, de autonomie van de school en de frequentie van examinering. Uit quantiele regressies (op basis van de TIMSS, PISA datasets) blijkt dat het effect van centrale eindexamens toeneemt met de bekwaamheid van de leerlingen, complementaire effecten heeft met de autonomie van de school en dat het effect toeneemt

⁷ TIMSS staat voor Third International Mathematics and Science Study

⁸ Opmerkelijk is verder het gebruik een variabele die het aantal boeken thuis aangeeft om te corrigeren voor de intellectuele achtergrond van de ouders.

gedurende de periode in het secundair onderwijs en dat reguliere gestandaardiseerde testen positieve effecten genereren. Er is dus sprake van een aanzienlijke heterogeniteit in de effecten van centrale examinering. Hierbij moet wel de kanttekening worden geplaatst dat verder onderzoek door middel van experimenten nodig blijft.

Uit de hier gepresenteerde literatuur komt naar voren dat centrale examinering positieve effecten heeft op de onderwijsprestaties. Over de optimale frequentie van examinering is echter niets gevonden.

7 Conclusies en Aanbevelingen

Deelnemers aan een huizenjacht zullen vaak het bekende mantra van de makelaar moeten aanhoren 'locatie, locatie, locatie'. Een degelijk credo zou niet misstaan bij onderzoek naar onderwijsinterventies waarbij de formulering uiteraard moet zijn: "causaliteit, causaliteit, causaliteit". In recente jaren is het aantal publicaties op basis van (quasi-) experimenteel onderzoek aanzienlijk toegenomen. Drie artikelen geven hiervan een mooi overzicht voor verschillende onderwijsinterventies (Webbink (2005), Van der Steeg (2005) en Cornet *et al* (2006)). Het hier gepresenteerde overzicht moet worden gezien als een update van deze inventarisaties.

Beleidsexperimenten

In Pomp (2003) wordt het uitvoeren van experimenten als leidraad voor beleid vergeleken met het uitvoeren van onzekere beleidsinitiatieven. Volgens hem is een belangrijk verschil tussen beide dat bij onzekere beleidsinitiatieven er allerlei mogelijkheden zijn om tegenvallende resultaten te verdoezelen ("de omstandigheden zijn veranderd", "zonder beleid was het nog erger geweest"). Maar dit is niet altijd noodzakelijk aangezien vaak onbekend is of een bepaalde maatregel wel een effect heeft gehad. Bekeken vanuit het perspectief van de beleidsmaker zijn experimenten minder attractief omdat 'Bij beleidsexperimenten komen tegenvallende resultaten duidelijk en voor iedereen zichtbaar aan het licht. De implicaties van een eventuele mislukking zijn dan ook onontkoombaar: het beleid dient te worden bijgesteld. Daarmee zijn beleidsexperimenten riskant voor beleidsmakers. Tegenvallende resultaten zullen bij buitenstaanders en belanghebbenden kritiek ontlokken in de trant van: dat was te voorzien, je had heel iets anders moeten doen etc.' Beleidsexperimenten vergen politieke moed aldus Pomp. De noodzaak hiervoor wordt ook ingegeven door het feit dat deze aanpak (afhankelijk van het type experiment) een lange adem vergt. Voor beleidsmakers kan dit leiden tot moeilijk uitlegbare situaties: het probleem is urgent en toch wordt in eerste instantie gekozen voor experimenten. Dit klinkt paradoxaal. Daar staat tegenover dat er eigenlijk geen alternatief is voor deze route: "Zonder beleidsexperimenten komen we nooit te weten wat werkt, en voeren we beleid met oogkleppen op. Met oogkleppen op naar de kenniseconomie: over paradoxen gesproken." (Pomp, 2003)

In de onderstaande tabel is schematisch de stand van zaken weergegeven. Afgezien van het thema 'schaalgrootte' is voor elk bekeken onderdeel literatuur te vinden waarbij door middel van experimenten de effecten van onderwijsinterventies is onderzocht. Voor het merendeel van de studies geldt echter dat zij niet in de Nederlandse setting hebben plaatsgevonden. Dit is geenszins een onoverkomelijk bezwaar en kan op z'n minst dienen als indicatie voor verder onderzoek.

Tabel 7-1 'Blinde' of 'witte' vlekken per thema

	Buitenland	Nederland
Onderwijspersoneel		
VSV/ leerwerkplicht		
VVE		
Schaalgrootte		
Examinering		

Voor twee van de vijf geselecteerde thema's bleek dat er geen publicaties zijn verschenen die voldoen aan de eis van (quasi-) experimenteel onderzoek én niet reeds behandeld in de drie overzichtsartikelen, te weten: 'voortijdig schoolverlaten/ leer-werkplicht' en 'schaalgrootte'. Ten aanzien van het laatste thema zijn in de afgelopen jaren wel enkele publicaties verschenen maar daarbij is het endogeniteitsprobleem niet expliciet aangepakt. In de tekst wordt een kort overzicht gegeven van de status van dit onderzoek. Voor het thema 'vsv/ leer-werkplicht' is een korte samenvatting opgesteld aan de hand van de drie overzichtsartikelen en de recente publicatie van het CPB "Voortijdig schoolverlaten in Nederland: omvang, beleid en resultaten" (Van der Steeg en Webbink, 2006).

In figuur 7-1 zijn de belangrijkste uitkomsten van de inventarisatie schematisch weergegeven. De uitkomsten voor het onderwijspersoneel worden, gezien de verschillende sub-onderdelen in de volgende sectie behandeld.

Figuur 7-1 Schematische weergave van uitkomsten op vier thema's



Ten aanzien van het VVE blijkt uit Nederlands en buitenlands onderzoek dat een vroege aanvang van onderwijs positieve effecten heeft op de (latere) onderwijsprestaties. Uit buitenlandse literatuur komt naar voren dat gestandaardiseerde testen (zoals centrale examinering) positieve

gevolgen hebben op de prestaties. Uit Amerikaans onderzoek blijkt dat *high stakes* testing (waarbij de uitslag directe gevolgen heeft voor de leerling in de vorm van doubleren of *summer school*) positieve effecten oplevert. Hierbij worden wel enkele kanttekeningen geplaatst. Bij het evalueren van een dergelijke aanpak dient ook te worden gekeken naar de mogelijke (onbedoelde) neveneffecten van een dergelijk systeem.

Onderwijspersoneel

Interventies ten aanzien van de kwaliteit en/of beloning van leraren is niet los te zien van de prikkels van andere stakeholders (leerlingen, ouders en directies). Dit is een belangrijk punt temeer omdat de vraag of (meer) concurrentie leidt tot betere prestaties van leerlingen nog niet is beslecht. In figuur 7-2 is de zogenaamde ‘prikkelketen’ weergegeven. Het laat de verschillende schakels zien alsmede de krachten die aanwezig moeten zijn willen bepaalde interventies succesvol zijn. Uit het onderzoek dat thans beschikbaar is blijkt dat ouders niet bijzonder gevoelig zijn voor openbare kwaliteitsindicatoren of in ieder geval hun beslissingen daar niet door laten leiden.⁹

Figuur 7-2 De prikkelketen voor interventies



Het huidige institutionele systeem bevat weinig prikkels ter verhoging van de kwaliteit van leraren. De beloning van leraren is vooral afhankelijk van de ervaring en in veel mindere mate van prestaties. In Cornet *et al* (2006) blijkt dat scholing en teambeloning van leraren kosteneffectiever zijn dan zijn bijvoorbeeld klassenverkleining. Tevens geven zij aan dat de kennis over de precieze determinanten nog beperkt is waardoor de mogelijkheid bestaat dat er betere alternatieven zijn die de kwaliteit van leraren verbeteren.

Kijkend naar de relatie tussen de kwaliteit van leraren en de kwaliteit van het onderwijs dan blijkt dat er geen gegevens voorhanden zijn voor de Nederlandse situatie. Dit is echter wel een gebied waarbij geldt dat kansrijke experimenten relatief eenvoudig kunnen worden opgezet. In het buitenland is wel ervaring opgedaan en daaruit blijkt dat de kwaliteit van de leraren vooral is terug te vinden op het vlak van taal (groter dan op rekenen) en dat ervaring de belangrijkste determinant is; zij-instromers hebben zelfs een negatief effect.

⁹ “Research into effects of the publications of public performance indicators is currently limited to research into the effects on choice behaviour of parents and pupils (and teachers in a few cases). The results of these studies are more unambiguous. They lead us, in any case, to draw the conclusion that, when choosing schools, parents and pupils take little notice of published performance indicators such as the school report card. Whether this takes place consciously or is due to unfamiliarity with the information on quality, is not known. On the other hand, principals and teachers have the opinion that performance indicators are important and there exists some evidence that accountability has a positive impact on student achievement. To put it differently, there exists no evidence that school report cards matter” (De Wolf en Janssens, 2005)

Voor het effect van de beloning van leraren op de kwaliteit van leraren geldt eveneens dat er geen publicaties zijn voor de Nederlandse situatie. Uit Amerika is een experiment bekend waarbij leraren dmv bonussen werden ‘gelokt’ naar scholen die ondermaats presteerden. Het effect hiervan was positief. Een gerelateerde vraag op dit onderwerp heeft te maken met de vraag op welke wijze de kwaliteit van leraren kan worden beoordeeld. Uit een onderzoek naar deze vraag blijkt dat het oordeel van het schoolhoofd/de directeur in vergelijking met schoolprestaties en oordelen van ouders een goed kompas is. Voor het derde subthema, de relatie tussen de beloning van leraren en de kwaliteit van onderwijs geldt wederom dat er geen specifiek Nederlands onderzoek is. Uit onderzoek in Chicago blijkt dat het zogenaamde *high stakes* testen positieve effecten kan hebben. De uitkomsten van een ‘bonustoernooi’ in Israël wijzen eveneens op een positieve impact van resultaatafhankelijke beloning op de kwaliteit van het onderwijs. Deze verandering bleek met name tot stand te komen door een verandering in onderwijsmethoden en de inzet van extra tijd voor onderwijs. Over de relatie tussen de rol van ondersteunend onderwijspersoneel en het maatschappelijk rendement van onderwijs is geen onderzoek voorhanden.

Enkele kansrijke experimenten voor de Nederlandse situatie

Het oplossen van het endogeniteitsprobleem in relatie tot onderwijsinterventies vraagt om (quasi-)experimenteel onderzoek. Uit de bovenstaande tekst blijkt dat er weinig dergelijke onderzoeken zijn uitgevoerd op Nederlandse bodem. Hieronder zijn enkele voorstellen voor experimenten opgenomen. Dit is slechts een kleine greep en dient dan ook niet meer als voorbeeld om te laten zien wat mogelijk is.

Voorstel voor evaluerend onderzoek: RMC-trajectbegeleiding

De RMC-functie heeft tot doel alle voortijdig schoolverlaters in beeld te krijgen en terug te leiden naar het onderwijs, of indien dit niet haalbaar is, te begeleiden naar een positie op de arbeidsmarkt. Het lijkt mogelijk om een evaluatie uit te voeren van de trajectbegeleiding van voortijdig schoolverlaters uit hoofde van de RMC-functie. Door capaciteitsproblemen bij RMC's krijgt een bepaalde groep voortijdig schoolverlaters geen behandeling. De arbeidsmarktpositie van deze groep kan worden vergeleken met die van de groep die wel wordt behandeld. Het voornemen is om in 2006 een vervolgonderzoek uit te voeren naar de effecten van de RMC-trajectbegeleiding. (afkomstig uit Van der Steeg en Webbink, 2006)

Onderzoeksvoorstel Evaluatie vsv-beleid: aansluiting van vmbo en het mbo

Het preventieproject “preventieproject overgang vmbo-mbo” heeft tot belangrijkste doel voortijdig schoolverlaten te voorkomen onder vmbo-gediplomeerden. Het project is in 2005 met ondersteuning van de Taskforce Jeugdwerkloosheid opgezet als pilot in vijf RMC-regio's. Kern van het project is zo vroeg mogelijk de risico-overstappers onder de vmbo-eindexamenleerlingen te herkennen. Vmbo-eindexamenleerlingen worden als risico-overstapper beschouwd als zij nog niet weten wat voor vervolgopleiding ze willen volgen, zich nog niet bij een ROC hebben aangemeld of een vervolgopleiding gekozen hebben waarvoor een wachtlijst geldt. Ook leerlingen die geen praktijkplaats hebben gevonden lopen risico omdat zij zonder die stageplek hun vervolgopleiding vaak niet kunnen starten. Ten slotte worden jongeren als risico-overstapper beschouwd als ze aangegeven hebben dat ze gaan werken en geen verdere scholing te willen

volgen. Onderdeel van het preventieproject is dat er contact wordt gelegd met deze leerlingen in de zomerperiode (juni-september). De begeleiding kan - afhankelijk van de risicofactor - bestaan uit telefonisch of schriftelijk contact om informatie te geven of leerlingen te volgen, een keuzebegeleidingsgesprek, nader onderzoek naar belangstelling of capaciteiten, of het verstrekken van schriftelijke informatie over opleidingen. Door de wijze waarop de pilot destijds is uitgerold is het mogelijk om deze introductie te bestuderen als een natuurlijk experiment. Het idee is om de vervolgbestemmingen van vmbo-gediplomeerden van deelnemende scholen aan het preventieproject te vergelijken met die van niet-deelnemende scholen. Hierdoor kunnen de werkelijke effecten van de interventie worden geïdentificeerd. (onderzoeksvorstel Van der Steeg en Webbink, CPB, 2006)

Het effect van invoering van de tweede fase (pakketprofielen & studiehuis)

In het jaarlijkse onderzoek Studie & Werk dat SEO Economisch Onderzoek uitvoert in opdracht van het weekblad Elsevier, worden in de meest recente peiling voor het eerst grote groepen van afgestudeerden waargenomen die in de bovenbouw van het voortgezet onderwijs te maken hadden met de invoering van de tweede fase. Met name onder afgestudeerde HBO'ers blijkt al een aanzienlijk deel (ongeveer 30%) eindexamen te hebben gedaan met een 'profielvakkenpakket'. Doordat de tweede fase niet in één jaar werd ingevoerd, maar scholen de vrijheid hadden om dat binnen enkele jaren te doen ontstaat de mogelijkheid om het effect van de tweede fase op studieprestaties in hoger onderwijs te schatten met een difference-in-differences schatter. De gedachte is dat cohorten van afgestudeerden voor de invoering van de tweede fase kunnen worden vergeleken met een cohort na invoering. Een onderzoek naar het effect van de tweede fase op studieprestatie in het hoger onderwijs kan snel en tegen lage kosten worden uitgevoerd.

Literatuur

- Bishop, John H. (2002): *What Is The Appropriate Role For Student Achievement Standards?*, in Kodrzycki, Yolanda K. (ed.) Conference Series 47 - Education in the 21st Century: Meeting the Challenges of a Changing World, p. 249-278, Federal Bank of Boston, Boston
- Clotfelter, Charles en Helen Ladd en Jacob Vigdor (2007): *How And Why Do Teacher Credentials Matter For Student Achievement?*, NBER working paper No. 12828, National Bureau of Economic Research, Cambridge MA, <http://www.nber.org/papers/w12828>
- Clotfelter, Charles en Elizabeth Glennie, Helen Ladd en Jacob Vigdor (2006): *Would Higher Salaries Keep Teachers In High-Poverty Schools? Evidence From A Policy Intervention In North Carolina*, NBER working paper No. 12285, National Bureau of Economic Research, Cambridge MA, <http://www.nber.org/papers/w12285>
- CPB (2006): Beoordeling KBA voortijdig schoolverlaten, CPB Notitie, juni 2006, Centraal Planbureau, Den Haag
- Currie, Janet en Matthew Neidell (2007): *Getting Inside the 'Black Box' of Head Start Quality: What Matters and What Doesn't*, Economics of Education Review, 2007, vol. 26, issue 1, pages 83-99
- Figlio, David N. en Lawrence Kenny (2006): *Individual Teacher Incentives And Student Performance*, NBER working paper No. 12627, National Bureau of Economic Research, Cambridge MA, <http://www.nber.org/papers/w12627>
- Gibbons, Stephen en Stephen Machin & Olmo Silva (2006): *Competition, Choice and Pupil Achievement*, Centre for the Economics of Education, London School of Economics
- Jacob, Brian A. (2005): *Accountability, Incentives and Behavior: Evidence from School Reform in Chicago*, Journal of Public Economics 89(5-6), 2005
- Jacob, Brian A. en Steven D. Levitt (2003): *Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating*, Quarterly Journal of Economics 118(3), 2003
- Jacob, Brian A. en L. Lefgren (2005): *Principals as Agents: Subjective Performance Assessment in Education*, KSG Working Paper No. RWP05-040, Harvard University - John F. Kennedy School of Government, <http://ssrn.com/abstract=757356>
- Jürges, Hendrik en Kerstin Schneider en Felix Büchel (2005): *The Effect Of Central Exit Examinations On Student Achievement: Quasi-Experimental Evidence From TIMSS Germany*, JEEA - September 2005, Vol. 3, No. 5, Pages 1134-1155
- Lavy, Victor (2005): *Performance Pay And Teachers' Effort, Productivity And Grading Ethics*, working paper, deze versie is afkomstig van <http://economics.huji.ac.il/faculty/lavy/papers.html>, eerder verschenen als NBER working paper No. 10622 (2004).
- Lavy, Victor (2006): *From Forced Busing To Free Choice In Public Schools: Individual And General Equilibrium Effects*, NBER Working Paper No. 11969, National Bureau of Economic Research, Cambridge MA, <http://www.nber.org/papers/W11969>

- Leigh, Andrew en Chris Ryan (2005): *Estimating Returns to Education: Three Natural Experiment Techniques Compared*, Discussion Paper No. 493, Centre for Economic Policy Research, Research School of Social Sciences, Australian National University
- Leuven, Edwin en Mikael Lindahl, Hessel Oosterbeek en Dinand Webbink (2006): *Expanding Schooling Opportunities for 4-Year-Olds*, IZA DP No. 2434, Forschungsinstitut zur Zukunft der Arbeit / Institute for the Study of Labor, Bonn
- Onderwijsraad (2005): *Variëteit in schaal, Keuzevrijheid, sociale samenhang en draagvlak bij grote onderwijsorganisaties*. Advies uitgebracht aan de minister van Onderwijs, Cultuur en Wetenschap, Onderwijsraad, Den Haag
- Pomp, Marc, (2003): *Bèta/techniek voor Innovatie en Productiviteit – Beleidsadvies*, SEO-rapport nr. 679, Stichting voor Economisch Onderzoek der Universiteit van Amsterdam, Amsterdam
- Steeg, M. van der (2005): *Why should governments intervene in education, and how effective is education policy? A review of the literature*, CPB Memorandum 122, Centraal Planbureau, Den Haag
- Steeg, M. van der en D. Webbink (2006): *Voortijdig schoolverlaten in Nederland: omvang, beleid en resultaten*, CPB Document No 107, Centraal Planbureau, Den Haag
- Venne, Louise van de (2002): *Schaal in het primair onderwijs. Een studie naar de relatie tussen schaal en organisatie-effectiviteit*, Amsterdam: Thela Thesis (dissertatie)
- Webbink, Dinand (2005): *Causal Effects in Education*, Journal of Economic Surveys Vol. 19, No. 4, p. 535-560
- Wolf, Inge F. de en Frans J.G. Janssens (2005): *Effects and side effects of inspections and accountability in education; an overview of empirical studies*, Working Paper no. 53, Amsterdam School of Economics, Amsterdam
- Wößmann, Ludger (2005): *The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-Repeat and PISA*, Education Economics, Volume 13, Number 2, June 2005, pp. 143-169(27)

Bijlage

In de onderstaande tabellen worden de belangrijkste uitkomsten weergegeven van (quasi-) experimenteel onderzoek: tabel A is een reproductie van tabel 5.1 uit Van der Steeg en Webbink (2006, p. 63), tabel B komt uit Webbink (2005), en tabel C is afkomstig van Van der Steeg (2005, p.25)

Tabel A Effecten van buitenlandse uitvalbestrijdingsprogramma's uit Van der Steeg en Webbink (2006)

Naam programma (land)	Omschrijving en doelgroep	Resultaten
School dropout demonstration assistance program (US)	Zestien verschillende programma's, waarvan acht gericht op 'middle school', en acht op 'high school'. Zowel organisatorische herstructurering van scholen (bv. kleinere scholen) als specifiek op achterstandsleerlingen gerichte projecten.	Over het algemeen geen vermindering uitval of verbetering studieresultaten; Intensieve 'middle school' programma's en GED-programma's voor high school leerlingen wel (gematigd) effectief.
Big brothers / big sisters (US)	Volwassen onbetaalde coach ('adult friend') voor leerlingen (10/16-jarigen) uit eenoudergezinnen.	Minder schoolverzuim; minder drugs- en alcoholgebruik en agressie; hogere cijfers.
Sponsor a scholar (US)	Intensieve langdurige begeleiding voor high school leerlingen door een mentor, financiële steun, hulp bij aanmelding voor college, academische steun.	Hogere gemiddelde cijfers; 16 (22) procent hogere deelname aan vervolgonderwijs ('college') na 1(2) jaar.
Quantum opportunity program (US)	Groepsmentor en financiële prikkels (twee dollar voor elk uur besteed aan verbetering van sociale vaardigheden of persoonlijke ontwikkeling) voor leerlingen uit lagere sociaal-economische milieus (4-jarig programma, 250 uur per jaar).	Hoger slagingspercentage voor high school of GED (26 procentpunt); hogere deelname aan college (13 procentpunt); minder crimineel gedrag (jongens zes keer zo vaak veroordeeld voor misdrijven).
Ohio's Learning, earning, and parenting (LEAP) program (US)	Case management en financiële prikkels (aanvulling op uitkering (20%) voor tienerouders voor deelname aan onderwijs. Korting op uitkering (20%) bij frequent verzuim of voortijdig schooluitval.	Hoger slagingspercentage voor high school en GED-programma; hoger salaris; lagere werkloosheid; echter geen positieve effecten indien al uitgevallen bij aanvang.
Education Maintenance Allowance (EMA) (UK)	Wekelijkse toelage en bonus gekoppeld aan prestaties (aanwezigheid, halen diploma) voor post-leerplichtigen (16-18 jarigen) die doorgaan met een voltijds opleiding (school of college), op voorwaarde dat ouders minder dan 30 000 pond per jaar verdienen en men zich aan het leercontract houdt.	Hogere participatie en behoud in postleerplichtig onderwijs (4 (6) procent hoger in eerste (tweede) jaar); grootste impact op zwakke en gemiddelde leerlingen en op jongens; geen significant effect op studieprestaties (o.a. cijfers).
High-school matriculation award (Israël)	Beloning (max. 1500 dollar) voor leerlingen van slecht presterende 'high schools' voor behalen van 'Bagrut'-certificaat. Extra beloningen voor behalen van tussentijdse toetsen en langere verblijfsduur.	Zes tot acht procent hoger slagingspercentage voor 'Bagrut' ('high school matriculation certificate', nodig voor toegang tot universiteit).
Teachers' performance incentive program (Israël)	Prestatiebeloning (gemiddeld 23 300 dollar per school) voor docenten van top 30% middelbare scholen, gemeten op basis van verbeteringen in prestaties van leerlingen (o.a. uitvalpercentage van de school en gemiddeld aantal studiepunten per student)	Hogere gemiddelde toetsscores; meer studiepunten per student (o.a. voor bètavakken); echter geen hoger percentage dat een 'Bagrut'-certificaat haalt, en ook geen lagere uitval. Positieve resultaten vooral voor zwakke leerlingen.
School resources program (Israël)	Extra middelen (gemiddeld 51 600 dollar) voor middelbare scholen met beste projectvoorstellen om schooluitval tegen te gaan en schoolprestaties te verbeteren. Middelen werden besteed aan extra lessen, kleinere klassen, training van leraren, en extra aandacht voor zwakke leerlingen.	Hogere gemiddelde testcores; meer studiepunten per student (niet voor bètavakken); geen hoger percentage dat een 'Bagrut'-certificaat haalt, en ook geen lagere uitval; positieve resultaten vooral voor beter presterende leerlingen.

Bronnen: Dynarski e.a. (1998), Heckman (2000), Lavy (2002), Angrist en Lavy (2004), Middleton e.a. (2004), Dearden e.a. (2005).

Bron: Van der Steeg en Webbink (2006, p. 63)

Tabel B Studies naar de effecten van onderwijsinterventies met een experimentele opzet, uit Webbink (2005)

Interventie	Oorsprong van exogene variatie	Effecten per type variatie	
		Alle variaties	Exogeen
Klasgrootte			
VS Krueger (1999)	STAR-experiment		+
Angrist and Lavy (1999)	Maximum class size rule	-/NS/+	+
Akerhielm (1995)	Average class size	-	+
Hoxby (2000a)	Natural randomness in size of population/ maximum/minimum class size	+	NS
Dobbelsteen et al. (2002)	Financing rule	NS	NS
Boozer and Rouse (1995)	Maximum class size special education	-	+
Case and Deaton, (1998)	Restriction by apartheid policy	NA	+
Lindahl (2001a)	Summer learning	-/ NS	+
Schooluren			
Lavy (1999)	Allocation rules school hours	-/ NS	+
Jacob and Lefgren (2004a)	Rules for promotion to the next grade	NS/+	NS/+
Lindahl (2001b)	Summer learning	NA	+
Training van onderwijzers			
Jacob and Lefgren (2004b)	Rules for probation of schools	NS	NS
Leuven et al., 2003)	Discontinuity in financing scheme	NA	NS
Computers			
Angrist and Lavy, (2002a)	Timing of delivery	-/ NS	-/ NS
Goolsbee and Guryan	Variation in subsidy rate	NA	NS
Leuven et al. (2003)	Discontinuity in financing scheme	NA	-
Peer groep			
Sacerdote (2001)	Random assignment dorm mates	NA	+
Zimmerman (2003)	Random assignment roommates	NA	+
Hoxby (2000b)	Natural randomness in gender and ethnicity of population	NA	+
Vouchers/schoolkeuze			
Rouse (2000)	Selection rule for enrolment	NA	+
Peterson et al. (2003) / Krueger and Zhu, 2002)	Voucher experiments		+
Cullen et al. (2003)	Lotteries for admission to schools	NA	NS
		NA	NS/+
Uitgaven per leerling			
Guryan (2001)	Discontinuities in equalization law allocating additional resources	NA	NS/+
Prestatiebeloning/ prikkels			
Angrist and Lavy (2002b)	Controlled experiment		NS/+
Leuven et al. (2003)	Controlled experiment		NS/+
Lavy (2002a)	Measurement error/discontinuity in assignment variable for treatment	NA	+
Lavy (2002a)	Selection rule into program	NA	+
Concurrentie			
Hoxby (2000c)	Location of great rivers	NA	+

–, significant negatief effect op resultaat; +, significant positief effect; NA, niet beschikbaar; NS, geen significant effect. – zie Webbink voor bronnen

Bron: Webbink (2005)

Tabel C Reproductie van tabel 3.1 uit van der Steeg (2005, p. 25)

Table 3.1 Effectiveness of public interventions in education by type and level (based on studies with an (quasi-) experimental approach)				
Timing (level)	Intervention (type) + country ^a	Literature	Effect Quality ⁷	Effect Quantity ⁷
Pre-Compulsory	early childhood interventions several pre-school intervention programs (e.g. full-day child care, home visits, pre-school program) (US)	Heckman (2000); Currie (2001) ⁸	+	+ ¹⁶
Compulsory ^d	'classical' education inputs Instruction time (school hours)			
	- Israel	Lavy (1999a)	+	
	- NL	Leuven et al. (2004a)	+ ⁹	
	- Sweden	Lindahl (2001)	+	
	expenditure per pupil (US)	Guryan (2001)	n.s./+ ⁹	
	class size reduction			
	- Israel	Angrist and Lavy (1999)	+	
	- NL	Dobbelsteen et al. (2002)	n.s.	
	- US	Krueger (1999, 2003)	+	
	- US	Hoxby (2000a)	n.s.	
	- US	Krueger and Whitmore (2001)	+	
	- various countries	Wößmann and West (2005)	n.s./+ ¹⁰	
	teacher training/acquisition			
	- Israel	Angrist and Lavy (2001)	n.s./+ ¹¹	
	- US	Jacob and Lefgren (2004a)	n.s.	
	teacher testing/certification (US)	Angrist and Guryan (2003)	n.s. ¹²	
	computer facilities			
	- In elementary and middle schools (Israel)	Angrist and Lavy (2002a)	n.s.	
	- In primary schools (NL)	Leuven et al. (2004b)	-	
	Internet investment subsidy	Goolsbee and Guryan (2002)	n.s.	
	organisational changes school-going age extension (Sweden)	Meghir and Palme (1999)		+
	competition (vouchers/school choice)			
	- US	Hoxby (2002)	+	
	- US	Cullen et al. (2003)	n.s.	
	- US	Krueger and Zhu (2003)	n.s.	
	performance incentives ^d			
	- merit pay for teachers (Israel)	Lavy (2003)	+	
	- merit pay for schools (Israel)	Lavy (2002)	+	n.s. ¹⁷
	peer group (changes in class heterogeneity)			
	- Israel	Lavy (1999b)	+	
	- US	Hoxby (2000b)	+	
	teachers' grading standards (US)	Figlio and Lucas (2000)	+	

Timing (level)	Intervention (type) + country ³	Literature	Effect Quality ⁷	Effect Quantity ⁷
	specific projects for students at risk			
	- cash bonus for high school matriculation for low-achieving students (Israel)	Angrist and Lavy (2002b)	+	
	- additional instruction to underperforming students (Israel)	Lavy and Schlosser (2004)	n.s./+ ¹³	
	- funding for extra personnel for primary schools with disadvantaged students	Leuven et al. (2004b)	n.s.	
	- student counselling and financial incentives for minority students (US)	Taggart (1995); Heckman (2000)	+	
	- several dropout prevention programs (US)	Dynarski et al. (1998) ⁶	n.s.	n.s.
	- remedial summer education (US)	Jacob and Lefgren (2004b)	n.s./+ ¹⁴	
Post-Compulsory ²				
	tuition fees			
	- US	Heckman et al. (1998),		-(0.07) ¹⁰
	- US	Cameron and Heckman (2001)		-(0.02-0.05) ¹³
	- US	Dynarski (1999)		-(0.03) ¹⁷
	financial support (loans/ grants)			
	- California college grant program (US)	Kane (2003)		+
	- social security student benefit (US)	Dynarski (1999)		+
	- SOFES loan program (Mexico)	Canton and Blom (2004)	+	
	performance incentives for students			
	- reward for 1st year college completion(NL)	Leuven et al. (2003)	n.s./+ ¹⁵	
	- performance-based grant system (NL)	Belot et al. (2004)	+	

Sources: partly adapted from Webbink and Hassink (2002); Webbink (2005).

1 Primary and secondary education.
2 Tertiary education.
3 US = United States, NL =Netherlands.
4 The reader is referred to Canton and Webbink (2004) for a discussion of evaluations of interventions based on performance incentives.
5 Heckman and Currie have reviewed several (quasi-)experimental evaluation studies of various early childhood intervention schemes.
6 Dynarski et al. (1998) have reviewed sixteen dropout prevention programs carried out in the US, of which eight middle-school programs and eight high-school programs. These programs were generally ineffective in lowering the dropout rate or improving student performance. However, some positive results were found for a limited number of intensive middle school programs and for high school GED programs.
7 + significant positive effect on performance; n.s. no effect; - significant negative effect on performance.
8 Positive effects on test scores apply only to pupils with lower educated parents and minority children.
9 Test scores are positively affected for 4th-grade students, but increased spending showed no effect on 8th-grade test scores
10 Class-size effects are estimated on mathematics and science achievement in 18 countries. Significant positive effects of smaller class size were found for only two countries (Iceland and Greece). Wößmann and West (2005) conclude that class-size effects in one school system cannot be interpreted as a general finding for all school systems.
11 Positive effects are found for secular schools, but effects on student performance in religious schools are insignificant.
12 Effects have been studied on teacher quality as an indirect measure of educational quality.
13 Remedial education improved the performance of sixth graders, but not of third graders.
14 Cash bonuses for high school matriculation (this is a prerequisite for university admission) are effective when provided to an entire school, but not when given to individually selected students within the school.
15 Positive effects are found for students with high math skills and for students with higher educated fathers when given a 'high' reward.
16 The quantity effect of early childhood interventions refers to the direct effect on average years of schooling due to the extra classes in the pre-school stage, but also to the indirect effect caused by increased high school graduation rates and/or lower dropout rates.
17 The quantity effect refers to the impact on the dropout rate.
18 Presented numbers are enrolment elasticities.



seo economisch onderzoek

Roetersstraat 29 . 1018 WB Amsterdam . T (+31) 20 525 16 30 . F (+31) 20 525 16 86 . www.seo.nl