

Sector : Groei, kennis en structuur  
Afdeling/Project : Leren van beleid  
Samenstellers : Dinand Webbink en Marc van der Steeg  
Nummer : 201  
Datum : 12 februari 2008

### **Essay voor de Parlementair Onderzoekscommissie Onderwijshervormingen**

#### **Samenvatting**

Beleid dat bijdraagt aan de kwaliteit van het Nederlands onderwijs kan hoge maatschappelijke baten opleveren. De effecten van beleid zijn echter vaak niet bekend. In de huidige beleidspraktijk blijven kansen voor evaluatie van beleid onbenut. Zo worden regelmatig ‘pilots’ gehouden of wordt beleid gefaseerd ingevoerd. Daardoor ontstaan automatisch groepen die wel en groepen die niet te maken krijgen met het nieuwe beleid. Door in dit soort situaties niet alleen de resultaten te volgen van de groepen die (als eerste) met beleid te maken krijgen, maar ook van soortgelijke groepen die er (nog) niet mee te maken hebben, wordt op eenvoudige wijze een wetenschappelijke experiment opgezet. Hiermee kan het effect van beleid worden vastgesteld. Het benutten van deze kansen voor beleidsevaluatie lijkt relatief weinig inspanningen te vergen, terwijl de baten aanzienlijk kunnen zijn.



# 1 Inleiding

Onderwijs is van cruciaal belang voor de Nederlandse samenleving en economie. Beleid dat het Nederlands onderwijs kan versterken kan daarom grote baten opleveren. Echter, de effecten van beleidsinterventies in het Nederlands onderwijs zijn meestal niet bekend. Zo concludeerde de Rekenkamer op basis van een overzicht van 35 evaluatiestudies dat er geen relatie kan worden gelegd tussen de ontwikkeling in prestaties van leerlingen met onderwijsachterstanden en het gevoerde beleid.<sup>1</sup> Maar ook op veel andere beleidsterreinen waar veel beleid is gevoerd en middelen zijn ingezet, bijvoorbeeld de bestrijding van voortijdig schoolverlaten of het invoeren van ICT in het onderwijs, is de kennis over de effecten van beleid gering. Nieuw beleid is meestal gebaseerd op kwalitatieve praktijkervaringen, meningen, opvattingen, en idealen, en niet op een harde empirische toetsing. Daaraan zijn grote risico's verbonden omdat de kosten van niet effectief of contraproductief beleid buitengewoon hoog kunnen zijn. Bezorgdheid over de effecten van recente grootschalige veranderingen in het onderwijs, zoals de invoering van de basisvorming en de Tweede Fase, ligt ook ten grondslag aan de beslissing van de Tweede Kamer om een parlementair onderzoek te starten naar deze recente onderwijshervormingen.

De laatste jaren is in binnen- (WRR, 2002, Onderwijsraad, 2006, CPB, 2004) en buitenland, met name in de Verenigde Staten<sup>2</sup> en door de OESO (2007) gepleit voor 'evidence-based policy'. Daarbij gaat het vooral om het vaker toepassen van experimenten ter onderbouwing van nieuw beleid. In dit essay, geschreven op verzoek van de Parlementaire Onderzoekscommissie 'Onderwijshervormingen', staat de vraag centraal hoe de wetenschappelijk onderbouwing van het onderwijsbeleid kan worden versterkt en meer kennis kan worden verworven over de effecten van beleid.

De ideale methode voor het vaststellen van de effecten van beleid is het klassieke gecontroleerde experiment. Het grote voordeel van gecontroleerde experimenten is dat hiermee overtuigende kennis kan worden verkregen over de effecten van beleid. Gezien het grote belang van 'human capital' in de huidige en toekomstige samenleving is deze kennis van grote waarde. Dit geldt nog sterker als het gaat om ingrijpende veranderingen zoals de invoering van basisvorming of de tweede fase in het voortgezet onderwijs.

Ondanks deze onomstreden voordelen stuiten beleidsexperimenten regelmatig op bezwaren. Experimenten zouden te lang duren, te duur zijn en ook ethische problemen opleveren vanwege de ongelijke behandeling van leerlingen. Echter, ook andere onderzoeksmethoden kennen vele problemen en leveren bovendien minder hard bewijs dan experimenten. Daarnaast vinden ook in de medische wetenschap regelmatig klassieke experimenten plaats. Daarbij worden zorgvuldige waarborgen gehanteerd voor deelname, voortzetting en stopzetting van het experiment.

<sup>1</sup> Algemene Rekenkamer (2001).

<sup>2</sup> Voor een overzicht, zie Onderwijsraad (2006), bijlage 4.

Overigens worden ook in de huidige beleidspraktijk leerlingen regelmatig ongelijk behandeld vanwege specifieke regelingen of 'pilots'. Deze ongelijke behandeling van leerlingen biedt mogelijkheden voor het vaststellen van de effecten van beleid. In feite doen zich in de huidige beleidspraktijk regelmatig situaties voor die vergelijkbaar zijn met experimenten, maar niet als zodanig worden herkend en daarom ook niet worden benut voor het verwerven van harde kennis. Bijvoorbeeld, in de huidige beleidspraktijk wordt veel gebruik gemaakt van 'pilots' en wordt beleid ook soms gefaseerd ingevoerd. Deze situaties kunnen worden benut om de effecten van beleid vast te stellen. De opzet en inrichting van beleid kunnen een eerste stap vormen voor het vaststellen van de effecten van beleid. Om te komen tot 'lerend beleid' is het derhalve belangrijk om bij de vormgeving van beleid aandacht te besteden aan de mogelijkheden om later de effecten te kunnen vaststellen.

Dit essay bespreekt mogelijkheden voor het vaststellen van beleidseffecten. Allereerst wordt ingegaan op de problemen bij het vaststellen van de effecten van beleid (paragraaf 2). Deze problemen kunnen worden opgelost door te zoeken naar adequate controlegroepen. Binnen de huidige beleidspraktijk zijn hiervoor verschillende mogelijkheden. In paragraaf 3 worden deze mogelijkheden besproken. De meest directe methode voor het vaststellen van de effecten van beleid is het opzetten en uitvoeren van gecontroleerde experimenten (paragraaf 3.1). Een tweede mogelijkheid is om bij de vormgeving van nieuw beleid meer aandacht te besteden aan kansen voor evaluatie (paragraaf 3.2). Paragraaf 4 sluit af met conclusies.

## 2 **Waarom weten we zo weinig over de effecten van beleid?**

In de huidige beleidspraktijk worden vele evaluaties uitgevoerd. Desondanks weten we nog weinig over de effecten van beleid. Dit komt in de eerste plaats omdat veel evaluaties zich richten op het proces van het invoeren van het nieuwe beleid en niet op het vaststellen van de effecten van beleid. Daarnaast worden vaak andere methoden gebruikt, zoals het verzamelen van oordelen van experts over de effectiviteit van beleidsmaatregelen. Dit zijn echter zwakke methoden voor het vaststellen van de effecten van beleid.

Voor een geloofwaardige evaluatie van de effecten van beleid zijn goede controlegroepen nodig die niet te maken hebben met het beleid. Het is echter vaak moeilijk om goede controlegroepen te vinden. Dit komt in de eerste plaats omdat beleid meestal over de hele linie wordt ingevoerd. In feite is dan alleen een vergelijking in de tijd mogelijk, voor en na invoering van de beleidsmaatregel. Het probleem is dan echter dat ook vele andere factoren kunnen veranderen waardoor niet duidelijk is of het gemeten effect veroorzaakt wordt door het nieuwe beleid of door andere factoren. Zo kunnen cohorten van leerlingen van elkaar verschillen en kunnen ook meerdere omgevingsfactoren veranderen. Een andere probleem is dat de invoering van nieuw beleid van invloed kan zijn op het gedrag van scholen en leerlingen. Bijvoorbeeld, bij de invoering van de Tweede Fase in Havo/Vwo daalde het aantal eindexamenkandidaten Havo van 21.000 in het laatste examenjaar Oude Stijl (2000) naar 14.000 in het eerste volledige examenjaar van de Tweede Fase (2001). Deze daling werd veroorzaakt door drie factoren: minder zittenblijven van leerlingen in Havo-4 bij het laatste cohort Oude Stijl, een hoger slagingspercentage voor het examen en minder doorstroom van Mavo gediplomeerden naar Havo na invoering van de Tweede Fase. Aannemelijk is dat de onzekerheid over de gevolgen van de invoering van de Tweede Fase scholen en leerlingen ertoe heeft aangezet om de laatste cohorten leerlingen zoveel mogelijk nog het oude examen te laten voltooien. Het gevolg hiervan is dat de samenstelling van het laatste cohort oude stijl Havo gediplomeerden waarschijnlijk sterk verschilt van het eerste cohort Tweede Fase Havo gediplomeerden. Een vergelijking van de latere resultaten van deze cohorten lijkt daarom dan ook weinig geschikt voor het vaststellen van het effect van de invoering van de Tweede Fase.

De tweede reden waarom het moeilijk is om goede controlegroepen te vinden is dat leerlingen, ouders en scholen voortdurend keuzes maken. Deze keuzes leiden tot (zelf)selectie van leerlingen tussen en binnen scholen. Leerlingen die wel te maken hebben met een bepaalde interventie, bijvoorbeeld kleine klassen, zullen door deze (zelf)selectie op veel factoren kunnen verschillen van leerlingen die niet te maken hebben met de interventie. Onderzoekers zullen veelal slechts rekening kunnen houden met een beperkt aantal factoren. Het gevolg is dat factoren die niet worden waargenomen, de resultaten kunnen vertekenen.

Stel dat een onderzoeker het effect van klassenverkleining op prestaties van leerlingen wil vaststellen. Zij kan hiervoor de leerprestaties vergelijken van leerlingen in kleine klassen met

die van leerlingen in grotere klassen en daarbij rekening houden met waargenomen verschillen tussen leerlingen, zoals leeftijd, geslacht en sociaaleconomische achtergrond. Echter, het effect dat op deze wijze wordt bepaald is vertekend als factoren die zowel van invloed zijn op de prestaties van leerlingen als op de selectie van leerlingen in kleine klassen niet worden waargenomen. Het is bijvoorbeeld mogelijk dat sterk gemotiveerde ouders kiezen voor scholen met kleine klassen en dat deze ouders hun kinderen ook stimuleren om zich in te spannen op school. Als de onderzoeker de motivatie van de ouders niet waarneemt kan zij het verschil in prestaties tussen leerlingen in kleine en grote klassen in dit geval ten onrechte toeschrijven aan de klassengrootte.

In de evaluatieliteratuur wordt dit probleem van selectie op grond van niet waargenomen kenmerken aangeduid als de endogeniteit van de interventie. Deelname van leerlingen aan een bepaalde interventie is niet aselekt (random) maar komt tot stand door (zelf)selectie op grond van waargenomen en niet waargenomen kenmerken. Vele studies van de afgelopen 15 jaar laten zien dat het onvoldoende rekening houden met selectie op grond van niet waargenomen factoren tot een sterke vertekening van de geschatte effecten van beleidsinterventies kan leiden. Dit probleem kan worden tegengegaan door te zoeken naar goede controlegroepen.

### **3 Mogelijkheden voor het vaststellen van effecten van beleid**

Om de effecten van beleid vast te stellen zouden we idealiter willen beschikken over een door loting tot stand gekomen experimentele groep, waarop het nieuwe beleid wordt toegepast, en een controlegroep, die niet te maken heeft met het nieuwe beleid. Cruciaal is dat leerlingen door toeval worden verdeeld over de experimentele en de controlegroep. Dit maakt het namelijk onwaarschijnlijk dat niet waargenomen factoren, zoals motivatie, de resultaten zullen vertekenen. De kans is immers heel klein dat door de loting bijvoorbeeld alle sterk gemotiveerde personen in de experimentele groep terecht komen en de niet gemotiveerde personen in de controlegroep. Het effect van het beleid kan worden bepaald door de resultaten van de experimentele groep te vergelijken met die van de controlegroep. Deze, vanuit evaluatie oogpunt, ideale situatie is (nog) vrij zeldzaam in het Nederlands onderwijsbeleid.<sup>3</sup> In paragraaf 3.1 bespreken we de toepassing van klassieke experimenten in het onderwijs. Belangrijk is ook dat zich binnen de huidige beleidspraktijk regelmatig situaties voor doen die niet ver verwijderd zijn van de ideale situatie van het klassieke experiment. Dit biedt kansen voor het overtuigend vaststellen van de effecten van beleid. Dit wordt besproken in paragraaf 3.2.

#### **3.1 Beleidsexperimenten**

De meest directe manier om hard bewijs te verkrijgen is het uitvoeren van klassieke gecontroleerde experimenten. In de Verenigde Staten zijn de afgelopen 20 jaar diverse experimenten uitgevoerd, bijvoorbeeld met programma's voor voor- en vroegschoolse educatie, het verkleinen van klassen, het toekennen van vouchers aan leerlingen met een lagere sociaaleconomisch achtergrond, en met de beloning van leraren.<sup>4</sup> Behalve in het onderwijs zijn in de Verenigde Staten ook op andere terreinen die ook relevant zijn voor het onderwijs, sociale experimenten uitgevoerd, bijvoorbeeld experimenten met het aanbieden van woonvouchers (Kling, et al. 2007).

Het grote voordeel van gecontroleerde experimenten is dat zij overtuigend empirisch bewijs bieden over wat werkt en niet werkt. Uiteindelijk vinden wetenschappers en beleidsmakers bewijs verkregen langs experimentele weg overtuigender dan bewijs dat op andere wijze is verkregen. Het is immers niet aannemelijk dat de resultaten beïnvloed worden door niet waargenomen factoren. Bijkomend voordeel is dat dit bewijs vaak helder gecommuniceerd kan worden.

<sup>3</sup> Het ministerie van OCW beschikt sinds kort over een experimenteerbudget waarmee de komende jaren enkele klassieke experimenten zullen worden ondersteund.

<sup>4</sup> Zie voor een overzicht Cornet en Webbink (2004).

Een nadeel van experimenteren is dat het opzetten en evalueren van een goed experiment niet eenvoudig is en kosten met zich brengt. Bovendien zijn de resultaten soms pas na enkele jaren bekend. Experimenteren kan een kwestie van lange adem zijn, terwijl de tijdshorizon van veel politici vaak beperkt is tot een kabinetsperiode. Echter, niet alle experimenten zijn kostbaar en langdurig. Dit hangt af van het type interventie dat wordt onderzocht. Zo is recent aan de Universiteit van Amsterdam een experiment uitgevoerd met financiële prikkels voor eerste jaar studenten.<sup>5</sup> Als deze studenten hun propedeuse zouden voltooien binnen een jaar zouden zij een financiële bonus ontvangen. Voor de controlegroep was er geen bonus. De resultaten van dit experiment waren na een jaar beschikbaar.<sup>6</sup> Momenteel worden experimenten uitgevoerd met de groepssamenstelling van studenten. Deze experimenten vergen weinig middelen en de resultaten zijn snel beschikbaar.

Een ander veelgenoemd nadeel van experimenteren betreft het ethische aspect. Een gecontroleerd experiment geeft per definitie ongelijke behandeling van mensen. Hier kan tegen worden ingebracht dat in de medische wetenschap experimenten met random toewijzing geheel geaccepteerd zijn. Waarom zou experimenteren in de sociale wetenschap dan ethisch niet verantwoord zijn? Ook is op voorhand niet zeker of de interventie de beoogde opbrengsten daadwerkelijk genereert. Verder staat tegenover het nadeel van ongelijke behandeling van een bepaalde groep leerlingen het voordeel dat toekomstige generaties kunnen profiteren van de nieuw verworven inzichten.

Om te bepalen of een experiment moet worden uitgevoerd dient een afweging gemaakt te worden tussen de kosten en baten van een experiment vergeleken met alternatieven. Vaak zal dit neerkomen op het beoordelen van de vraag of de grotere betrouwbaarheid van experimentele resultaten opweegt tegen de extra kosten. Vooral bij ingrijpende en omvangrijke beleidswijzigingen lijkt de waarde van overtuigend bewijs uit een gecontroleerd experiment groot ten opzichte van de kosten van zo'n experiment. Daarbij speelt ook dat de kosten van het voeren en continueren van niet-effectief beleid zeer hoog kunnen zijn.

## **3.2 Kansen voor effectmeting binnen de huidige beleidspraktijk**

Binnen de huidige beleidspraktijk doen zich regelmatig kansen voor om effecten van beleid overtuigend vast te stellen. Voor het benutten van deze kansen is de vormgeving van het beleid cruciaal. Hieronder bespreken we enkele kansen die tot dusver nog niet worden benut.

### **3.2.1 Pilots benutten voor effectmeting**

Op veel beleidsterreinen wordt gebruik gemaakt van 'pilots'. Vaak wordt dan het nieuwe beleid toegepast bij een aantal scholen of instellingen die hiervoor belangstelling hebben. Het doel is

<sup>5</sup> Leuven et al., 2003.

<sup>6</sup> De financiële prikkel bleek de studieprestaties te verbeteren van de studenten met de hoogste eindexamencijfers.



vaak om de eerste kinderziekten van het nieuwe beleid op te sporen. ‘Pilots’ worden evenwel niet benut om de effecten van het nieuwe beleid vast te stellen. Als echter ‘pilots’ worden uitgebreid met controlegroepen is dit wel mogelijk. Allereerst zouden scholen kunnen worden gevraagd of zij belangstelling hebben om mee te doen aan de pilot. Vervolgens zou de pilot door loting kunnen worden toegewezen aan een aantal scholen. De resultaten kunnen worden gemeten op beide groepen scholen en gebruikt om het effect van het nieuwe beleid vast te stellen. Daarmee wordt een stap gezet van een pilot naar een gecontroleerd experiment. Door hier systematisch aandacht aan te besteden kan belangrijke winst worden geboekt in de richting van ‘evidence based policy’.

### **3.2.2 Gefaseerde invoering van beleid**

Nieuw beleid wordt vaak in fasen ingevoerd. Sommige scholen starten eerder en de rest volgt na enkele jaren. Bijvoorbeeld, bij de invoering van de Tweede Fase werd in 1998 gestart met een eerste groep scholen en de rest volgde in 1999. Ander voorbeeld, eind jaren negentig werden omvangrijke middelen beschikbaar gesteld voor ICT in het onderwijs. Volgens het actieplan ‘Investeren in voorsprong’ zou gestart worden met een aantal voorhoedescholen. Daarna kwamen andere scholen gefaseerd aan bod. Dergelijk gefaseerd beleid biedt mogelijkheden voor het vaststellen van de effecten omdat sommige scholen en leerlingen wel te maken hebben met het nieuwe beleid en andere scholen en leerlingen (nog) niet. Door een vergelijking van de resultaten op ‘vroeg starters’ met degenen die nog niet zijn begonnen wordt het mogelijk om effecten van beleid te bepalen. Om gefaseerd beleid echt te benutten is het belangrijk dat de vroeg starters door heldere criteria worden bepaald. Anders gezegd, selectie in de groep ‘vroeg starters’ op grond van niet waargenomen factoren mag geen rol spelen. Het gebruik maken van gefaseerd beleid is in feite een grootschalig ‘pilot’. Overwogen kan worden om bij omvangrijke beleidswijzigingen, waarvan de uitkomsten ongewis zijn, te starten in bepaalde regio’s. De resultaten en ervaringen in deze regio’s kunnen dan gebruikt worden voor de besluitvorming over het al of niet opschalen van dit beleid.

### **3.2.3 Aselect toekennen van schaarse middelen**

Bij veel subsidies zijn de middelen niet toereikend voor alle aanvragers. Daardoor zullen sommige aanvragers geen subsidie ontvangen. De toekenning van subsidies kan worden gebruikt om een controlegroep te maken. Een voorbeeld kan worden ontleend aan het Amerikaanse Head Start project. In dit project worden middelen voor voorschoolse educatie beschikbaar gesteld aan bepaalde doelgroepen, met name leerlingen uit arme gezinnen. Recent deed een adviescommissie voor de evaluatie van Head Start de aanbeveling om, in gevallen dat er te weinig middelen beschikbaar zijn voor alle deelnemers, de toewijzing van leerlingen door loting te laten plaatsvinden. Daarmee wordt voor de toekomst een overtuigende evaluatie van het effect van het programma mogelijk gemaakt.

### 3.2.4 Gebruik maken van discontinuïteiten in regelingen

Het onderwijsbeleid kent diverse regelingen waarin sprake is van afkapgrenzen. Scholen boven een bepaald criterium ontvangen wel additionele middelen en scholen net onder het criterium komen daarvoor niet in aanmerking. Een vergelijking van de resultaten van scholen net boven de afkapgrens met die van scholen net onder de afkapgrens kan een overtuigende schatting van het effect van het beleid opleveren. Een recent voorbeeld betreft de evaluatie van een maatregel waarmee extra geld voor computers en personeel werd toegekend aan scholen met minstens 70 % achterstandsleerlingen.<sup>7</sup> Scholen net beneden deze grens kregen geen extra subsidie, scholen vanaf 70 % kregen wel extra subsidie. Door scholen rond de grens van 70 procent te vergelijken (en daarbij rekening te houden met het percentage achterstandsleerlingen) wordt een gecontroleerd experiment nagebootst. Scholen met minstens 70 procent achterstandsleerlingen vormen de experimentele groep, scholen net beneden de grens van 70 procent vormen de controlegroep. Een ander voorbeeld betreft het verlenen van beurzen voor het volgen van een studie in het buitenland in het kader van het Talentenprogramma. Voor het toekennen van deze beurzen werd door een commissie een ranking gemaakt van de kwaliteit van de aanvragers. Aan de hand van deze ranking kon een controlegroep gemaakt worden waarmee het effect van studeren in het buitenland is gemeten (Oosterbeek en Webbink, 2005).

Een mogelijkheid is om bewust verschillen (discontinuïteiten) in subsidieregelingen in te bouwen, analoog aan het hiervoor besproken voorbeeld van de extra subsidie voor scholen met minstens 70 procent achterstandsleerlingen. Door in de regelingen enige verschillen in het beschikbaar stellen van middelen in te bouwen ontstaan mogelijkheden om de effecten vast te stellen.

<sup>7</sup> Zie Leuven e.a. (2007).

## 4 Conclusies

Kennis over de effecten van beleid kan hoge baten opleveren. Het is echter vaak niet eenvoudig om kennis te verwerven over de effecten van beleid. De belangrijkste reden is dat beleid meestal over de hele linie wordt ingevoerd waardoor er geen adequate controlegroepen beschikbaar zijn. De meest direct manier om de effectiviteit van beleid vast te stellen is het uitvoeren van gecontroleerde experimenten. Binnen de huidige beleidspraktijk blijven kansen voor evaluatie onbenut. Verschillende situaties kunnen worden uitgebouwd tot een gecontroleerd experiment of tot een vergelijkbare opzet voor evaluatie van beleid. Dit kan door:

- Pilots uit te breiden met controlegroepen;
- Beleid gefaseerd in te voeren, bijvoorbeeld starten in bepaalde regio's;
- Bij het toekennen van subsidies controlegroepen te maken van aanvragers die geen subsidie ontvingen;
- Afkapgrenzen (discontinuïteiten) in regelingen vaker te benutten. Controlegroepen kunnen worden gevormd uit personen of scholen die net niet in aanmerking komen voor de subsidies.

Meer kennis over de effecten van beleid heeft waarschijnlijk hoge maatschappelijke baten, vooral bij grootschalige ingrepen in het onderwijs. Het daadwerkelijk realiseren van 'lerend beleid', dat wil zeggen beleid waarin systematisch aandacht bestaat voor het meten van de effecten van beleid, vereist 'commitment' van beleidsmakers. Allereerst is geduld nodig, omdat kennis verwerven over de effecten van beleid tijd vergt. In de tweede plaats is 'commitment' nodig om de vormgeving van beleid ook daadwerkelijk aan te passen, zodat toekomstige evaluaties van hoge kwaliteit mogelijk zijn. Afstemming tussen beleidsmakers en evaluatieonderzoekers bij de implementatie van beleid is hierbij belangrijk.

## Referenties

Algemene Rekenkamer, 2001, *Bestrijding van onderwijsachterstanden*, Den Haag.

CPB, 2002, *De pijlers onder de kenniseconomie: opties voor institutionele vernieuwing*, Centraal Planbureau, Den Haag.

Cornet, M.F. en H.D. Webbink, 2004, Lerend beleid: het versterken van beleid door experimenteren en evalueren, CPB Document no. 48.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, Experimental Analysis of Neighborhood Effects, *Econometrica* 75:1 (January 2007), 83-119.

Leuven, E., M. Lindahl, H. Oosterbeek and H.D. Webbink, 2007, The effect of extra funding for disadvantaged students on achievement. *Review of Economics and Statistics*, 89 , (4), 721-736.

Leuven, E., H. Oosterbeek and B. van der Klaauw, 2003, *The effect of financial rewards on students achievement: Evidence from a randomized experiment*, University of Amsterdam.

OECD, 2007, *Evidence in education: Linking research and policy*, Paris.

Onderwijsraad, 2006, *Naar meer evidence based onderwijs*, Den Haag.

WRR, 2002, *Van oude kennis en nieuwe kennis, De gevolgen van ICT voor het kennisbeleid*, Den Haag.