

CPB Document

No 187

Juli 2009

Wat is bekend over de effecten van kenmerken van onderwijsstelsels? Een literatuurstudie

Dinand Webbink, Inge de Wolf (Inspectie van het Onderwijs),
Ludger Woessmann (Universiteit van München), Roel van Elk,
Bert Minne en Marc van der Steeg

Centraal Planbureau
Van Stolkweg 14
Postbus 80510
2508 GM Den Haag

Telefoon (070) 338 33 80
Telefax (070) 338 33 50
Internet www.cpb.nl

ISBN 978-90-5833-412-1

Korte samenvatting

Deze studie bespreekt de empirische evidentie over de effecten van kenmerken van onderwijsstelsels in het realiseren van de publieke belangen: kwaliteit, doelmatigheid en toegankelijkheid. Voor het onderzoek zijn drie stappen uitgevoerd. Allereerst is een selectie gemaakt van relevante stelselkenmerken op basis van een vergelijking van Nederland met een aantal relevante referentielanden. Vervolgens is voor elk stelselkenmerk een overzicht gemaakt van de internationale kennis over de effecten van deze kenmerken op de publieke belangen. Daarbij is de nadruk gelegd op de meest geloofwaardige studies, vooral studies met een (quasi-) experimentele opzet. Ook is gebruik gemaakt van recente studies waarin stelselkenmerken van landen gerelateerd worden aan prestaties in internationale toetsen. In de derde stap zijn de bevindingen vertaald naar een onderzoeksagenda.

Steekwoorden: kenmerken van onderwijsstelsels, kwaliteit, doelmatigheid, toegankelijkheid

Abstract

This study reviews the empirical evidence on the effects of characteristics of educational systems on public interests, like the quality of education, efficiency and accessibility. First, a selection of relevant characteristics has been made by comparing the Dutch system with several other countries. Next, for each characteristic a review has been made of the available evidence. The review focuses on the results from the most credible studies, especially studies that (quasi-) use experimental designs. In addition, recent cross-country studies that relate characteristics of educational systems with student achievements in international tests are also used. Finally, the main findings have been translated in a research agenda.

Key words: characteristics of education systems, quality, efficiency, accessibility

A comprehensive summary is available from www.cpb.nl.

Inhoud

Ten geleide	7
Samenvatting	9
1 Introductie	21
2 Aanpak	23
3 Stelselkenmerken	27
3.1 Concurrentie, vrije school keuze en segregatie	27
3.2 Autonomie en verantwoording	34
3.3 Selectie	40
3.4 Leraren en schoolleiders	43
3.5 Speciaal onderwijs	54
3.6 Achterstandenbeleid	57
3.7 Excellentie	66
3.8 School en omgeving	69
4 Onderzoeksagenda	71
Referenties	75

Ten geleide

Op verzoek van de Programmaraad voor het Onderwijsonderzoek (PROO) heeft het CPB een overzichtsstudie gemaakt naar vraagstukken op het terrein van het onderwijsbestel. Het verzoek van de PROO richtte zich op de volgende vragen:

- Wat is bekend over de effectiviteit van onderwijssystemen in het realiseren van publieke belangen?
- Wat is er bekend over de weging van publieke belangen in het onderwijsbeleid?

De publieke belangen zijn de kwaliteit, doelmatigheid en toegankelijkheid van het onderwijs. Ook verzocht de PROO om aandacht te besteden aan de aard en de kwaliteit van de ‘evidence’ alsmede lacunes en witte vlekken in de ‘evidence’ te identificeren en in kaart te brengen wat de relevante vraagstellingen zijn op dit terrein. Het voorliggende rapport geeft een overzicht van de empirische kennis over kenmerken van het Nederlandse onderwijsstelsel.

Het rapport is geschreven door Inge de Wolf (Inspectie van het Onderwijs), Ludger Woessman (Universiteit van München) en CPB-medewerkers Roel van Elk, Bert Minne, Marc van der Steeg en Dinand Webbink. Wij danken Kasper Weekenborg, Sander Baljé, Daisy Satijn, André de Moor, Ruud Okker en George Gelauff voor commentaar op een eerdere versie van het rapport. Tot slot danken wij de PROO voor de financiële ondersteuning van dit project.

Coen Teulings
Directeur CPB

Samenvatting

Op verzoek van de Programmaraad voor het Onderwijsonderzoek (PROO) heeft het CPB een overzichtsstudie gemaakt naar vraagstukken op het terrein van het onderwijsbestel. Het verzoek van de PROO richtte zich op de volgende vragen:

- Wat is bekend over de effectiviteit van onderwijssystemen in het realiseren van publieke belangen?
- Wat is er bekend over de weging van publieke belangen in het onderwijsbeleid?

De publieke belangen zijn de kwaliteit, doelmatigheid en toegankelijkheid van het onderwijs. Ook verzocht de PROO om aandacht te besteden aan de aard en de kwaliteit van de ‘evidence’ alsmede lacunes en witte vlekken in de ‘evidence’ te identificeren en in kaart te brengen wat de relevante vraagstellingen zijn op dit terrein.

Het onderzoek is uitgevoerd in drie stappen. Allereerst is een selectie gemaakt van relevante stelselkenmerken. Door PROO is een aantal kenmerken genoemd. Deze lijst is uitgebreid met markante verschillen van het Nederlandse onderwijsstelsel met enkele relevante referentielanden. Vervolgens is voor elk stelselkenmerk een overzicht gemaakt van de internationale kennis over de effecten van deze kenmerken op de publieke belangen. Daarbij is de nadruk gelegd op de meest geloofwaardige studies, dat zijn de kwalitatief beste studies op basis van methodologische criteria. Ook is gebruik gemaakt van recente studies waarin stelselkenmerken van landen gerelateerd worden aan prestaties in internationale toetsen, zoals PISA en TIMMS. In de derde stap zijn de bevindingen vertaald naar onderzoeksvragen. Voor elk stelselkenmerk zijn de bevindingen hieronder samengevat.

Concurrentie, vrije schoolkeuze en segregatie

In het Nederlandse onderwijs hebben leerlingen/ouders de mogelijkheid om zelf een school te kiezen. Landen zoals de Verenigde Staten of het Verenigd Koninkrijk kennen restricties bij de keuze van de school. Zo wordt in de Verenigde Staten nog heftig gedebatteerd over de introductie van vrije schoolkeuze. Vanuit internationaal perspectief kan het Nederlandse stelsel worden gekarakteriseerd als een stelsel met vrije schoolkeuze. De stelselvragen die in Nederland spelen gaan vooral over het bevorderen dan wel verminderen van de schoolkeuze. De mogelijkheden voor schoolkeuze kunnen worden bevorderd door het opheffen van restricties op het toetreden van scholen of het aanbieden van opleidingen. De vrije schoolkeuze wordt verminderd door het introduceren van restricties, zoals het postcodebeleid in het Amsterdamse basisonderwijs.

Vrije schoolkeuze kan leiden tot concurrentie tussen scholen en ook tot segregatie van bepaalde groepen leerlingen. Concurrentie kan scholen stimuleren de kwaliteit van het

onderwijs te verbeteren. De conclusie uit het empirisch onderzoek lijkt te zijn dat concurrentie bijdraagt aan de prestaties van leerlingen. Een vrije schoolkeuze biedt leerlingen en ouders de mogelijkheid om de school te kiezen die het best aansluit bij hun wensen en behoeften. Een betere 'match' tussen leerling en school kan de prestaties van leerlingen verbeteren. Het empirisch onderzoek laat zien dat de prestaties van leerlingen met een lagere sociaaleconomische achtergrond toenemen als zij een kwalitatief betere school bezoeken. Ook blijkt dat landen met meer schoolkeuze betere scores op internationale toetsen. De mogelijke keerzijde van een vrije schoolkeuze is dat hierdoor meer ongelijkheid kan ontstaan, bijvoorbeeld als leerlingen met een lagere sociaaleconomische achtergrond vaker kiezen voor scholen met een lagere kwaliteit. Recent onderzoek voor de Verenigde Staten bevestigt dit: gezinnen met lagere inkomens kiezen vaker voor scholen met een lagere kwaliteit, gemeten in testcores. Het aanbieden van transparante informatie over schoolkwaliteit blijkt invloed te hebben op de schoolkeuze in deze gezinnen. Bovendien blijken de overstappers betere resultaten te halen op scholen met een hogere kwaliteit. Vrije schoolkeuze kan ook leiden tot segregatie. Concentratie van leerlingen met een lage sociaaleconomische achtergrond op bepaalde scholen kan nadelig zijn vanwege zogenoemde 'peereffecten'. Het empirisch onderzoek levert bewijs voor het optreden van peereffecten, maar deze effecten lijken klein te zijn.

Wat betekent deze empirische evidentie voor het Nederlandse stelsel? Hoewel niet alle resultaten eenduidig zijn, lijkt het algemene beeld uit de literatuur te zijn dat concurrentie en vrije schoolkeuze leiden tot gemiddeld betere prestaties, en derhalve positief bijdragen aan *de kwaliteit van het onderwijs als publiek belang*. Daar staat echter tegenover dat de ongelijkheid in prestaties kan toenemen door vrije schoolkeuze. Als vrije schoolkeuze ertoe leidt dat de sociaaleconomische achtergrond van een leerling samenhangt met de kwaliteit van de school kan gesproken worden van een vermindering van *de toegankelijkheid van het onderwijs*. Het bevorderen van concurrentie en keuzemogelijkheden, bijvoorbeeld door de toetreding van scholen te bevorderen, kan leiden tot betere prestaties. Dit kan echter ook meer ongelijkheid in de hand werken. Hier lijkt derhalve sprake van een uitruil tussen kwaliteit en ongelijkheid. De ongelijkheid kan worden verminderd door het beschikbaar stellen van eenvoudig toegankelijke informatie. Het verschil in keuzegedrag naar sociaaleconomische achtergrond in relatie tot de beschikbaarheid van transparante informatie over kwaliteit van scholen lijkt belangrijk voor het Nederlandse onderwijsstelsel en roept de volgende vragen op:

- Kiezen ouders met een lagere sociaaleconomische achtergrond in Nederland vaker voor scholen met een lagere kwaliteit?
- In hoeverre hangt dit samen met de transparantie en toegankelijkheid van informatie over de kwaliteit van scholen in Nederland?
- In hoeverre leiden verschillen in schoolkeuze naar sociaaleconomische achtergrond tot verschillen in prestatieprikkels voor scholen?

- Zijn scholen met een goede kwaliteit terughoudend in het verspreiden van informatie over hun kwaliteit, om daarmee te voorkomen dat leerlingen van slecht geïnformeerde ouders voor hun school kiezen?

Autonomie en verantwoording

Vanuit internationaal perspectief kennen Nederlandse scholen een grote mate van autonomie. Autonomie is direct verbonden met verantwoording en toezicht, bijvoorbeeld door het afnemen van centrale examens, het beoordelen van prestaties van scholen of door de Onderwijsinspectie. Empirisch onderzoek laat zien dat de prestaties van leerlingen beter zijn in landen waarin scholen autonomie hebben over het onderwijsproces en over personele beslissingen. Centrale examens blijken een sterk positief effect te hebben op de prestaties van leerlingen. Dit roept de vraag op of het gebruik van schoolexamens in het Nederlandse voortgezet onderwijs, de positieve effecten van centrale examens vermindert. In de VS worden sinds enkele jaren verschillende zogenoemde ‘school accountability’ systemen gebruikt. Dit zijn systemen waarin de prestaties van scholen worden beoordeeld volgens van tevoren vastgelegde standaarden en waarin consequenties worden verbonden aan het niet halen van de normen. Evaluaties van deze systemen wijzen op positieve effecten maar laten ook zien dat strategisch gedrag van scholen veel voorkomt. Een goed design van dit soort systemen is derhalve van groot belang. Welke aanpak voor ‘falende of slecht presterende scholen’ het beste werkt is niet bekend.

Wat betekent deze evidentie voor het Nederlandse stelsel? Het empirisch onderzoek laat zien dat autonomie van scholen en verantwoording door centrale examens of ‘school accountability systemen’ positief bijdragen aan *de kwaliteit van het onderwijs*. Strategisch gedrag van scholen is echter een ‘fact of life’, en kan nadelig zijn voor *de toegankelijkheid* van het onderwijs als scholen bepaalde leerlingen weren. Voor de onderzoeksagenda lijken vooral vragen aan de orde over:

1. Het ontwerp van goede beoordelingssystemen van scholen;
2. De aanpak van falende of slecht presterende scholen.

Selectie

De selectie in het Nederland onderwijs vindt vooral plaats bij de overgangen tussen onderwijstypen. In de eerste plaats gaat het dan om de overgang van primair naar voortgezet onderwijs. Na het basisonderwijs stromen leerlingen door naar verschillende niveaus van voortgezet onderwijs. Andere landen, zoals de Verenigde Staten of de Scandinavische landen, selecteren pas op latere leeftijd. Het Nederlandse onderwijs kent derhalve een vroege selectie van leerlingen. Andere overgangen doen zich voor na het afronden van verschillende vormen van voortgezet onderwijs. De internationale empirische literatuur kijkt vooral naar de invloed van een vroege selectie van leerlingen. Er bestaat geen eenduidig beeld van de invloed van vroege selectie op de kwaliteit van het onderwijs. Wel lijkt het empirisch onderzoek er op te

wijzen dat vroege selectie van leerlingen de invloed van de sociaaleconomische achtergrond van leerlingen op hun onderwijsuitkomsten versterkt. Anders gezegd, vroege selectie lijkt te leiden tot meer ongelijkheid in onderwijskansen, en vermindert derhalve *de toegankelijkheid*. Voor Nederland is gevonden dat uitstel van selectie, voor leerlingen met een Mavo-advies, kan leiden tot meer hoger opgeleiden. De empirische literatuur geeft geen antwoord op de vraag in hoeverre latere doorstroommogelijkheden eventuele nadelige effecten van vroege selectie ongedaan kunnen maken.

Voor Nederland lijkt de vraag aan de orde hoe mogelijke nadelige effecten van vroege selectie het best kunnen worden tegengegaan, hetzij door uitstel van selectie van leerlingen, hetzij door het vergroten van latere doorstroommogelijkheden.

Leraren en schoolleiders

Leraren spelen een sleutelrol in het onderwijs. Empirisch onderzoek laat zien dat de kwaliteit van leraren een sterk effect heeft op de prestaties van leerlingen. Verbetering van de kwaliteit van docenten kan daarom leiden tot betere prestaties van leerlingen. De empirische literatuur biedt echter weinig houvast over de determinanten van de kwaliteit van leraren. De invloed van de initiële opleiding van de docent op de prestaties van leerlingen is niet duidelijk, empirische studies geven geen consistent beeld. Ook de prestaties die leraren zelf haalden tijdens hun onderwijsloopbaan zijn geen goede voorspeller van hun bijdrage aan de prestaties van leerlingen. Training van werkende leraren en begeleiding van startende leraren door een mentor kan wel effect hebben. De ervaring van de docent blijkt slechts beperkt van invloed te zijn op de prestaties. Alleen docenten in de eerste twee jaren van hun beroepsloopbaan dragen minder bij aan de prestaties van leerlingen dan docenten met meer ervaring. Na twee jaar in het onderwijs blijkt ervaring nog maar weinig verschil te maken. De beloningsstructuur in het onderwijs is vooral gebaseerd op ervaring en veel minder op prestaties. In verschillende landen wordt geëxperimenteerd met verschillende vormen van prestatiebeloning voor leraren. Deze experimenten wijzen erop dat het relateren van de beloning van leraren aan prestaties positief bijdraagt aan de prestaties van leerlingen. In een recente studie zijn ook positieve effecten gevonden van prestatiebeloning voor schoolleiders. Het empirisch onderzoek over leraren gaat vooral over de invloed op *de kwaliteit van het onderwijs* en levert weinig inzicht in de effecten op *de toegankelijkheid* en *de doelmatigheid*. Wel lijkt prestatiebeloning in het algemeen doelmatiger dan het verhogen van salarissen.

Voor het Nederlandse onderwijsstelsel roept dit empirische onderzoek de volgende vragen op:

1. Allereerst spelen er vragen over de aansluiting tussen opleiding en beroep. De empirische literatuur vindt geen sterke relatie tussen de initiële opleiding van de docent en de prestaties van leerlingen. Wat betekent dit voor de werving van kwalitatief goede leraren en voor de inrichting

- van de lerarenopleiding? Zijn opleidingseisen nodig of vormen ze juist een obstakel voor het aantrekken van goed personeel door het beperken van het aanbod van potentiële kandidaten?
2. De beloning van leraren is vooral gebaseerd op ervaring. Ervaring blijkt echter alleen in de eerste twee jaren van de beroepsloopbaan van invloed te zijn op de prestaties. Het relateren van de beloning van leraren aan de prestaties van leerlingen blijkt de kwaliteit van het onderwijs te verhogen. Wat zijn de mogelijkheden voor en de effecten van prestatiebeloning in Nederland?

Speciaal onderwijs

Vrijwel alle landen kennen naast het reguliere onderwijs vormen van speciaal onderwijs voor leerlingen met een bepaalde handicap. Daarbij kunnen verschillen optreden ten aanzien van de verhouding regulier/speciaal (wat is het aandeel leerlingen speciaal onderwijs binnen de totale leerlingenpopulatie) en de toewijzing aan het speciaal onderwijs. Ook worden leerlingen uit het speciaal onderwijs soms opgevangen of herplaatst in het reguliere onderwijs. De empirische kennis over deze stelselkenmerken en hun effecten op *kwaliteit, doelmatigheid en toegankelijkheid* is erg beperkt. Tussen landen bestaat aanzienlijke variatie in aandelen leerlingen in het speciaal onderwijs. Het effect daarvan op de totale onderwijsuitkomsten is niet bekend. De empirische literatuur geeft aanwijzingen dat de vraag naar speciaal onderwijs gevoelig is voor de financieringsystematiek. Voor de VS is het oorzakelijk effect geschat van een programma van het speciaal onderwijs. Het programma bleek positief bij te dragen aan de prestaties in rekenen en in taal. De effecten van het opvangen van leerlingen uit het speciaal onderwijs in het regulier onderwijs (voor henzelf en hun klasgenoten) zijn niet bekend. Mede gezien de substantieel hogere kosten van het speciaal onderwijs vergeleken met het regulier onderwijs lijkt verder onderzoek naar de effectiviteit van verschillende vormen van speciaal onderwijs in Nederland, alsmede van het opvangen van deze leerlingen in het regulier onderwijs, zeer gewenst.

Achterstandenbeleid

Nederland kent verschillende vormen van achterstandenbeleid. Zo zijn in de afgelopen jaren meer middelen ingezet voor programma's voor vóór- en vroegschoolse educatie (VVE) met als doel het verminderen van onderwijsachterstanden bij risicogroepen. Daarnaast ontvangen scholen in het Nederlandse basisonderwijs extra middelen voor leerlingen met een lagere sociaaleconomische achtergrond. Ook worden de laatste jaren meer middelen ingezet om het voortijdig schoolverlaten tegen te gaan.

Voor kwalitatief goede VVE-programma's zijn in de Verenigde Staten positieve effecten gevonden bij risicoleerlingen. Evaluaties op basis van gecontroleerd experimenten hebben laten zien dat deze daadwerkelijk geleid hebben tot een verbetering van het onderwijsniveau en de latere arbeidsmarktpositie van de deelnemers. Daarbij bleek dat er ook substantiële baten waren voor de rest van de samenleving in termen van vermindering van criminaliteit en inactiviteit. De vormgeving van de projecten bleek belangrijk te zijn. Intensieve projecten van minimaal vier

dagdelen per week met goed opgeleide begeleiders lijken betere resultaten te geven dan minder intensieve programma's.

In het Nederlandse onderwijs, vooral in het basisonderwijs, ontvangen scholen meer middelen voor leerlingen met een lagere sociaaleconomische achtergrond. Deze zogenoemde compenserende bekostiging is niet geëvalueerd in onderzoek met een experimentele opzet. Empirisch onderzoek naar de effecten van het verhogen van de onderwijsmiddelen van scholen, zonder daaraan prestatie-eisen te verbinden, geeft een wisselend beeld. In een recente Nederlandse evaluatie, met een experimentele opzet, werd gevonden dat achterstandsscholen niet beter gingen presteren nadat zij 10% meer middelen hadden ontvangen. In Nederland worden achterstandsmiddelen vaak gebruikt voor het verkleinen van klassen. De empirische literatuur laat zien dat kleinere klassen wel leiden tot betere prestaties maar dat dit zeer veel middelen vergt. Anders gezegd, klassenverkleining leidt wel tot meer *kwaliteit* van onderwijs maar is niet *doelmatig*.

Voortijdig schoolverlaten is een hardnekkig probleem. De empirische literatuur laat zien dat veel programma's niet werken. Daarnaast blijkt de effectiviteit van interventies vaak niet bekend te zijn. Studies met een experimentele opzet (inclusief controlegroepen) geven twee veelbelovende richtingen. In de eerste plaats blijken projecten die gebruik maken van financiële prikkels voor leerlingen, leraren en scholen effectief te zijn. Positieve ervaringen zijn hiermee opgedaan in het Verenigd Koninkrijk, de VS, en Israël. Ook zijn veelbelovende ervaringen opgedaan met langdurige en intensieve programma's met coaches, gericht op de sociale ontwikkeling van risicjongeren.

Een algemene les uit de empirische literatuur over vermindering van onderwijsachterstanden lijkt te zijn dat vroeg ingrijpen het best werkt: voorkomen werkt beter dan genezen. Effectieve interventies voor risicogroepen dragen zowel bij aan *de kwaliteit van het onderwijs als aan de toegankelijkheid van het onderwijs*. De inzet van achterstandsmiddelen in Nederland, die vooral gericht is op VVE en het primair onderwijs, lijkt in lijn met de empirische evidentie. De vragen die nu aan de orde zijn gaan over:

1. Het vinden van de meest effectieve programma's en interventies in de Nederlandse context, aanpakken die kosteneffectiever zijn dan klassenverkleining.
2. Het prikkelen van scholen tot het gebruik van de meest kosteneffectieve aanpakken. Andere belangrijke vragen, die ook eerder naar voren kwamen, gaan over schoolkeuze en segregatie, en daaraan verbonden vragen over de kwaliteit van (leraren op) scholen met veel achterstandsleerlingen.

Excellentie

Een vergelijking van de prestaties van Nederlandse leerlingen met leerlingen uit andere rijke landen laat zien dat Nederland goed scoort aan de linkerkant van de verdeling van kennis en vaardigheden, maar minder goed scoort aan de rechterkant van de verdeling (Minne et al.

2007). Anders gezegd, qua excellentie lijkt er ruimte voor verbetering te zijn voor het Nederlandse onderwijssysteem. Recent onderzoek levert aanwijzingen dat excellentie in het onderwijs de welvaart bevordert zowel voor het individu als voor de samenleving als geheel. Over het bevorderen van excellentie is echter weinig bekend. De empirische literatuur levert weinig inzicht in welk beleid hiervoor het meest effectief is. Financiële prikkels voor studenten blijken de prestaties van de betere leerlingen te bevorderen (maar geen effect te hebben op de overige leerlingen). Een toename van private middelen voor het hoger onderwijs biedt de mogelijkheid om de kwaliteit van opleidingen te verhogen. De invloed van het verhogen van private bijdragen op de toegankelijkheid van het hoger onderwijs, lijkt op basis van Amerikaanse studies, klein te zijn. In een recente Amerikaanse studie zijn echter ook aanwijzingen voor leenangst gevonden.

Voor het Nederlandse onderwijs lijkt de vraag aan de orde te zijn op welke wijze de excellentie het best kan worden bevorderd. Een eerste vraag is waar nieuw beleid het best bij kan aangrijpen: bij hoger onderwijs, voorgezet onderwijs of meteen al bij het basisonderwijs? Nader onderzoek kan zich ook richten op het effect van recent geïntroduceerde programma's voor excellente leerlingen in het voortgezet en hoger onderwijs. Ook zijn er vragen rond de inrichting (selectie aan de poort) en financiering van het hoger onderwijs (differentiatie van collegegelden).

School en omgeving

De laatste jaren worden scholen steeds meer verbonden met andere actoren of instellingen uit de omgeving van de school. Een voorbeeld hiervan is de zogenoemde 'Brede school', waarin de school bijvoorbeeld samenwerkt met culturele en sportinstellingen en er ook na de reguliere schooltijd activiteiten zijn voor leerlingen of kinderen uit de buurt. Onderzoek naar deze activiteiten suggereert dat leerlingen op deze scholen beter presteren. Het gaat hierbij echter niet om studies met een experimentele opzet. Het is daarom niet duidelijk of de 'brede school' daadwerkelijk leidt tot positieve uitkomsten omdat de uitkomsten vertekend kunnen zijn door de selectie van ouders en scholen. Een opvallende ontwikkeling in de omgeving van het Nederlandse publieke onderwijs is de groei van huiswerkbegeleiding door private aanbieders. Er bestaan geen experimentele studies naar de effecten hiervan op de prestaties van leerlingen.

Voor het Nederlandse onderwijsstelsel lijken daarom vragen aan de orde over de effecten van 'brede scholen' en van huiswerkbegeleiding door private aanbieders.

Onderzoeksagenda

Als we het Nederlandse onderwijsstelsel afzetten tegen de empirische evidentie blijkt dat de opzet van het Nederlandse stelsel op verschillende punten als, in potentie, effectief kan worden gekarakteriseerd. Nederland kent al geruime tijd een vrije schoolkeuze en concurrentie tussen scholen. Nederlandse scholen hebben een grote mate van autonomie. Centrale eindexamens zorgen voor transparantie over de prestaties van leerlingen en scholen. Daarnaast bewaakt de

Onderwijsinspectie de kwaliteit van het onderwijs. Ook volgt de aanpak van achterstanden in het onderwijs op hoofdlijnen de meest kansrijke benadering volgens de empirische evidentie. Voor- en vroegschoolse programma's worden steeds meer ingezet en veel achterstandsmiddelen worden ingezet bij de jongste leerlingen. Het is echter niet duidelijk of deze, in potentie, sterke elementen van het Nederlandse onderwijs, volledig tot hun recht komen en daadwerkelijk hun doel bereiken. Hierbij speelt de vormgeving van beleid een belangrijke rol. Vaak is niet duidelijk of de gekozen aanpakken en interventies daadwerkelijk effectief zijn.

Transparantie informatie over de kwaliteit van scholen

Vanuit internationaal perspectief heeft het Nederlandse onderwijsstelsel een aantal opvallende kenmerken: vrije schoolkeuze, concurrentie tussen scholen en een grote mate van autonomie voor scholen. Voor een goede werking van een dergelijk stelsel is transparante informatie over de kwaliteit (prestaties) van scholen van groot belang. Hiervoor zijn twee redenen. Allereerst stelt een goede informatievoorziening ouders in staat om een goede school te kiezen. In de tweede plaats is transparantie informatie over prestaties van scholen belangrijk als tegenwicht voor de autonomie van scholen.

Een vrije schoolkeuze is gebaat bij een goede informatievoorziening over de kwaliteit van scholen en bij een goed gebruik van deze informatie bij de schoolkeuze. Wanneer slechts een deel van de ouders de beschikbare informatie gebruikt zal vrije schoolkeuze maar beperkt bijdragen aan de concurrentie. Als dit gebruik van informatie sterk varieert naar sociaaleconomisch achtergrond kan vrije schoolkeuze bovendien zorgen voor meer ongelijkheid naar sociaaleconomische achtergrond. De recente plannen in de gemeente Nijmegen, om via een centraal meldpunt te komen tot een andere spreiding van leerlingen over scholen, lijken ook ingegeven door bezorgdheid over de uitkomsten van de vrije schoolkeuze. Daarom lijkt de vraag aan de orde in hoeverre de schoolkeuze in Nederland leidt tot gewenste resultaten, en in hoeverre de informatievoorziening daaraan bijdraagt of aanpassing behoeft. Recent onderzoek in de Verenigde Staten laat zien dat eenvoudig toegankelijke en transparante informatie over de kwaliteit van scholen invloed heeft op het keuzegedrag van ouders en leerlingen, en dat overstappers beter presteren op de nieuw gekozen scholen. Onderzoek naar de rol van informatie bij de schoolkeuze in Nederland lijkt daarom gewenst.

Het Nederlandse onderwijs kent een grote mate van autonomie van scholen. Transparante informatie over de kwaliteit van scholen werkt disciplinerend en stimuleert scholen tot goede prestaties. Het ontwerp van goede beoordelingssystemen van scholen is daarom belangrijk voor Nederland. Daarnaast spelen vragen over hoe om te gaan met slecht presterende scholen. Falende scholen lijken de achilleshiel van het Nederlandse stelsel. Over de omgang met falende scholen bestaat nauwelijks harde evidentie. Wel zijn er aanwijzingen dat zwak presterende scholen in staat zijn zich te verbeteren. Hoe dit het best gerealiseerd kan worden, door ondersteuning of door prikkels en sancties, is niet duidelijk. Onderzoek naar beste aanpak voor falende of slecht presterende scholen lijkt daarom belangrijk voor het Nederlandse onderwijs.

De kwaliteit van leraren en schoolleiders

Leraren spelen een sleutelrol in het onderwijs. Een cruciale vraag betreft het bevorderen van de kwaliteit van leraren. In het bijzonder spelen er vragen over de aansluiting tussen opleiding en beroep. De empirische literatuur vindt geen sterke relatie tussen enerzijds de initiële opleiding of de toetsscores van de docent en anderzijds de prestaties van leerlingen. Wat betekent dit voor de werving van kwalitatief goede leraren? Zijn opleidingseisen nodig of vormen ze juist een obstakel voor het aantrekken van goed personeel doordat ze het potentiële aanbod verminderen? Een andere belangrijke vraag betreft de beloning van leraren. De beloning van leraren en schoolleiders is vooral gebaseerd op ervaring. De empirische literatuur leert dat ervaring nauwelijks van invloed is op de prestaties van leraren. Alleen in de eerste twee jaren heeft het ontbreken van ervaring een negatief effect. Recente studies vinden positieve effecten van het gebruik van prestatiebeloning in het onderwijs. Dit roept de vraag op naar de mogelijkheden om de beloning van leraren en schoolleiders te relateren aan hun prestaties. Wat zijn de mogelijkheden voor en de effecten van prestatiebeloning in het Nederlandse onderwijs?

Centrale examens versus schoolexamens

De empirische literatuur geeft stevig bewijs voor positieve effecten van het gebruik van centrale examens op de kwaliteit van het onderwijs. Dit roept de vraag op of het gebruik van schoolexamens, naast centrale examens, nadelig is voor de kwaliteit van het onderwijs.

Vroege selectie

Een opvallend aspect aan het Nederlandse onderwijs is de vroege selectie van leerlingen na de basisschool. Recent onderzoek suggereert dat vroege selectie voor bepaalde groepen leerlingen de kansen vermindert op het behalen van een hoger onderwijs diploma. Een belangrijke vraag is op welke wijze de eventuele nadelige effecten van vroege selectie van bepaalde groepen leerlingen kunnen worden tegengegaan, hetzij door uitstel van selectie van leerlingen, hetzij door het vergroten van latere doorstroommogelijkheden.

Effectieve interventies

Op verschillende terreinen zijn vragen aan de orde over de effectiviteit van de gekozen aanpakken en interventies. Zo is over de effectiviteit van de stelselkenmerken van het speciaal onderwijs weinig bekend. Nederland kent de laatste jaren een verschuiving van leerlingen naar duurdere vormen van speciaal onderwijs. Wat zijn de effecten van deze vormen van speciaal onderwijs voor de betreffende leerlingen maar ook voor klasgenoten bij opvang in het regulier onderwijs? Bij de aanpak van onderwijsachterstanden in Nederland is het de vraag welke interventies het meest kosteneffectief zijn, en hoe scholen kunnen worden gestimuleerd om deze aanpakken ook daadwerkelijk te gebruiken? Excellentie in het onderwijs is een relatief nieuw onderwerp op de beleidsagenda. Ook hier spelen vragen over de effectiviteit van beleid gericht op het bevorderen van excellentie. Waar kan beleid het best aangrijpen: bij hoger

onderwijs, voorgezet onderwijs of meteen al bij het basisonderwijs? Nader onderzoek kan zich ook richten op het effect van recent geïntroduceerde programma's voor excellente leerlingen in het voortgezet en hoger onderwijs. Tot slot, is over de effecten van de brede school of van huiswerkbegeleiding door private instellingen weinig bekend.

Welk type onderzoek?

Het vaststellen van de effecten van kenmerken van het onderwijsstelsel is lastig omdat in het onderwijs tal van selectieprocessen een rol spelen. Ouders en leerlingen kiezen scholen, daarnaast maken scholen voortdurend keuzen met betrekking tot het leerproces en de inzet van docenten. Veranderingen in het onderwijsstelsel kunnen grote effecten hebben en brengen derhalve risico's met zich. Deze risico's kunnen worden verkleind door beleidsbeslissingen te onderbouwen met kwalitatief hoogwaardig onderzoek. Dit betreft onderzoek waarin rekening wordt gehouden met selectiviteit door gebruik te maken van geloofwaardige onderzoeksdesigns, zoals experimenten en natuurlijke experimenten.

Opvallend aan deze overzichtsstudie is dat deze weinig Nederlandse studies met een (quasi)experimenteel design bevat. Het werken met controlegroepen en voor- en nametingen is nog relatief nieuw in het Nederlandse onderwijsonderzoek, en random toewijzing is een uitzondering. Vanuit de onderzoeks-, beleids- en onderwijswereld worden vaak diverse redenen aangedragen om niet de hoogste methodologische eisen te stellen. Een vergelijkbare situatie bestond enkele jaren geleden in Amerika.¹ Inmiddels is dit type onderzoek daar wel doorgebroken. In Nederland gaan binnenkort ook experimenten van start ondersteund door middelen vanuit het actieprogramma Onderwijsbewijs². Het bevorderen van dit type onderzoek, dat solide bewijs oplevert voor beleidsbeslissingen, kan hoge maatschappelijke opbrengsten genereren. De kosten van slecht onderbouwd beleid, dat na verloop van tijd niet blijkt te werken, zijn namelijk zeer hoog.

Lerend beleid

Nieuw beleid is de afgelopen decennia in het Nederlandse onderwijs vaak over de hele linie ingevoerd. Als dit beleid niet werkt kan de schade groot zijn. Bovendien is het lastig om de effecten vast te stellen omdat er geen controlegroepen zijn als beleid over de hele linie wordt ingevoerd. Anders gezegd, er zijn grote risico's verbonden aan het invoeren van nieuw beleid over de gehele linie. Deze risico's kunnen worden verminderd door nieuw beleid eerst op kleinere schaal in te voeren. Als daarbij gekozen wordt voor een opzet met een experimenteel design kan solide bewijs worden verkregen over de effecten van het nieuwe beleid. Dit bewijs kan vervolgens worden gebruikt bij de beslissing om het nieuwe beleid geheel in te voeren, dan wel bij te stellen of stop te zetten. Mogelijkheden voor dit lerend beleid doen zich voor bij beleid dat gefaseerd wordt ingevoerd, zoals de eerste ronde van het convenantenbeleid ten

¹ Zie Cook (2002) voor een bespreking van discussie over experimenteel onderzoek in het onderwijs.

² <http://www.minocw.nl/onderwijsbewijs/index.html>

aanzien van het voortijdig schoolverlaten. Een andere mogelijkheid is om bij pilotprojecten gebruik te maken van controlegroepen. Vaak zijn er onvoldoende middelen om alle leerlingen aan een nieuwe interventie te kunnen laten deelnemen. In dat soort situaties kan door loting worden bepaald wie wel en wie niet aan de nieuwe interventie gaan deelnemen. Door te loten heeft iedereen evenveel kans. Het benutten van dit soort mogelijkheden vereist een goede afstemming tussen beleid en onderzoek. De meeste onderdelen van de hierboven geschetste onderzoeksagenda lenen zich voor een dergelijke aanpak.

1 Introductie

Het onderwijsbestel is het geheel van instituties, regels en organisaties in het onderwijs. Het onderwijsbestel dient een aantal publieke doelstellingen, zoals de kwaliteit van het onderwijs, de doelmatigheid van de besteding van middelen voor onderwijs en de toegankelijkheid van het onderwijs. Op verzoek van de Programmaraad voor het Onderwijsonderzoek (PROO) is een overzichtsstudie gemaakt naar vraagstukken op het terrein van het onderwijsbestel. Het verzoek van de PROO richtte zich op de volgende vragen:

- Wat is bekend over de effectiviteit van onderwijssystemen in het realiseren van publieke belangen?
- Wat is er bekend over de weging van publieke belangen in het onderwijsbeleid?

Ook verzocht de PROO om aandacht te besteden aan de aard en de kwaliteit van de ‘evidence’ alsmede lacunes en witte vlekken in de ‘evidence’ te identificeren en in kaart te brengen wat de relevante vraagstellingen zijn op dit terrein.

Het voorliggende rapport bevat de resultaten van de overzichtsstudie. De opbouw van het rapport is als volgt. In paragraaf 2 wordt de aanpak geschetst. Het gaat daarbij vooral om de selectie van relevante stelselkenmerken en de selectie van studies uit de literatuur. Paragraaf 3 geeft voor elk stelselkenmerk een overzicht van de internationale kennis. In paragraaf 4 worden de bevindingen vertaald naar een onderzoeksagenda.

2 Aanpak

Het onderzoek is uitgevoerd in drie stappen. Allereerst is een selectie gemaakt van relevante stelselkenmerken. Voor deze selectie is het Nederlandse onderwijsstelsel als uitgangspunt genomen en op hoofdlijnen vergeleken met het stelsel van een aantal relevante referentielanden. De kenmerken kunnen betrekking hebben op alle typen onderwijs in Nederland; van voor- en vroegschoolse educatie tot en met het hoger onderwijs. Vervolgens is voor elk stelselkenmerk een overzicht gemaakt van de internationale kennis over de effecten van deze kenmerken op de publieke belangen. Daarbij is de nadruk gelegd op de meest geloofwaardige studies, dat zijn de kwalitatief beste studies op basis van methodologische criteria. Ook is gebruik gemaakt van recente studies waarin stelselkenmerken van landen gerelateerd worden aan prestaties in internationale toetsen zoals PISA en TIMMS. In de derde stap zijn de bevindingen vertaald naar onderzoeksvragen.

Publieke belangen

Het doel van het project is het in kaart brengen van de effecten van kenmerken van onderwijsstelsels op publieke belangen. Daarbij gaat het om de effecten op kwaliteit, doelmatigheid en toegankelijkheid. In de literatuur worden verschillende uitkomstmaten gehanteerd die gerelateerd zijn aan deze publieke belangen. Het overgrote deel van de studies richt zich op de effecten van stelselkenmerken op toetscores (en soms ook lonen), als maat voor de kwaliteit van het onderwijs³. Een veel kleiner deel van de studies gaat ook in op de verhouding tussen effecten en kosten (kosteneffectiviteit), waarmee zicht wordt verkregen op de doelmatigheid. De toegankelijkheid van het onderwijs wordt meestal onderzocht door te kijken naar verschillen in deelname voor leerlingen met een verschillende sociaaleconomische achtergrond, bijvoorbeeld de deelname aan het hoger onderwijs of aan verschillende niveaus van voortgezet onderwijs. In dit project zullen deze uitkomstmaten steeds gebruikt worden als benadering voor het effect op publieke belangen. In het project zal ook worden gekeken naar mogelijke uitruilen (trade-offs) tussen publieke belangen. Stelselkenmerken kunnen namelijk van invloed zijn op meerdere publieke belangen. Bijvoorbeeld, een verhoging van de kwaliteit door selectie aan de poort van het hoger onderwijs kan ten koste gaan van de toegankelijkheid van het hoger onderwijs. Recente studies wijzen erop dat een hogere kwaliteit belangrijk is voor economische groei (Hanushek & Woessman, 2008, Minne et. al. 2007). Daar staat echter een grotere ongelijkheid tegenover. In het overzicht zullen dit soort trade-offs worden besproken.

³ We zijn ons ervan bewust dat de opbrengsten van onderwijs breder zijn dan de gemeten testcores of lonen. Deze studie geeft echter een overzicht van de beschikbare literatuur en die kijkt vooral naar testcores en lonen.

Selectie van kenmerken van onderwijsstelsels

Onderwijsstelsels zijn veelvormig en hebben vele kenmerken. Tussen landen bestaan grote verschillen in opzet, organisatie en financiering van het onderwijs. Er zijn bijvoorbeeld verschillen in de aanvangsleeftijd en de duur van het leerplichtig onderwijs, de overgangen tussen primair en voortgezet onderwijs, de mogelijkheden voor ouders om een school te kiezen, en de aansturing van scholen. Bij deze veelheid aan verschillen tussen landen is het de vraag welke kenmerken te selecteren. In de aanvraag voor de overzichtsstudie worden door de PROO de volgende kenmerken genoemd: aanwezigheid centrale examens, mate van autonomie, het beloningssysteem, schooltypen en klassensamenstelling, en selectie en differentiatie. In deze studie is deze lijst nog met een aantal kenmerken uitgebreid. Deze uitbreidingen betreffen vooral markante verschillen van het Nederlandse onderwijsstelsel met die van enkele relevante referentielanden (de 'Nordic countries' (Zweden, Noorwegen, Finland), Angelsaksische landen (VS en VK) en buurlanden (Duitsland, België en Frankrijk)). Zo kent Nederland bijvoorbeeld een vrije schoolkeuze en veel aandacht voor kinderen met leerachterstanden. Bij de selectie speelde ook mee welke kenmerken van het Nederlandse stelsel regelmatig in beleidsdiscussies aan de orde zijn. Tot slot speelde bij de selectie ook de vraag mee of de kenmerken zich redelijkerwijs lenen voor beleidsaanpassingen. Deze aanpak leidde tot de volgende selectie van kenmerken/thema's (met subkenmerken):

- Vrijheid van schoolkeuze
 - Concurrentie
 - Schoolkeuze en informatie
 - Segregatie;
- Autonomie en verantwoording
 - Autonome scholen
 - Verantwoording (centrale eindexamens, accountability systemen, inspectie toezicht, aanpak falende scholen)
- Overgangen in het stelsel (vroeg selectie)
- Leraren en schoolleiders
- Speciaal onderwijs
- Achterstandenbeleid
 - Voor- en voerschoolse educatie
 - Compenserende bekostiging
 - Voortijdig schoolverlaten
- Excellentie
 - Opbrengsten van excellentie
 - Prikkels
 - Private bijdragen

- School en omgeving.

Selectie van studies

Vervolgens is voor elk kenmerk/thema een overzicht gemaakt van de internationale evidentie over de effecten op de publieke belangen. De literatuur over de effecten van onderwijsstelsels is buitengewoon omvangrijk en beslaat meerdere disciplines. Voor het maken van een overzicht is het noodzakelijk om studies te selecteren. Een belangrijke vraag is dan welke studies geselecteerd worden en op grond van welke criteria. In dit onderzoek zijn studies geselecteerd op basis van methodologische criteria. Studies zijn ingedeeld op basis van hardheid. Het meest hard zijn studies volgens de zogenoemde gouden standaard: klassieke gecontroleerde experimenten met aselechte toewijzing aan de experimentele en controle groep. De zilveren standaard betreft studies die ook gebruik maken van een aselechte controlegroep, maar deze is tot stand gekomen door een toevallige situatie in de werkelijkheid (natuurlijke experimenten, bijvoorbeeld regressiediscontinuïteiten, Leuven et al. 2007). De bronzen standaard betreft studies die door middel van een specifiek onderzoeksdesign trachten om niet geobserveerde verschillen tussen individuen of groepen te elimineren. Bijvoorbeeld door gebruik te maken van een controlegroep met voor en nameting (difference-in-differences methode). In dit overzicht is de nadruk gelegd op de evidentie afkomstig van de meest geloofwaardige studies, dat zijn de kwalitatief beste studies volgens de gouden, zilveren of bronzen standaard. Studies waarin alleen rekening wordt gehouden met geobserveerde verschillen tussen groepen of individuen, door het opnemen van zoveel mogelijk controlevariabelen, worden minder geloofwaardig geacht omdat de resultaten vertekend kunnen zijn door niet geobserveerde verschillen. In onderwijsonderzoek speelt dit een belangrijke rol omdat selectie door ouders en scholen voortdurend plaats vindt in het onderwijs. Daarnaast wordt in deze studie een overzicht gemaakt van de internationale literatuur volgens twee lijnen. De eerste lijn betreft de literatuur die zich richt op het vergelijken van onderwijsstelsels tussen landen of staten binnen landen. Recent zijn veel studies verschenen waarin de effectiviteit van stelselkenmerken wordt onderzocht aan de hand van de resultaten op internationale toetsen, zoals PISA en TIMSS (zie Woessmann, 2006, 2007, Schütz et al. 2008, Hanushek & Woessmann, 2008). De tweede lijn betreft de literatuur over de effectiviteit van kenmerken van onderwijsstelsels op basis van studies binnen landen. In het algemeen is de geloofwaardigheid van het laatste type studies groter omdat het in studies tussen landen erg lastig blijkt te zijn om te controleren voor alle verschillen tussen landen. Ook is het de vraag of de resultaten te generaliseren zijn vanwege de vele verschillen tussen landen. Echter, voor een aantal kenmerken blijkt het lastig om variatie te vinden binnen landen. In dat geval kunnen studies tussen landen indicaties geven, hoewel de interpretatie van de resultaten van deze studies enige terughoudendheid vereist.

Naar een onderzoeksagenda

In het laatste deel van dit onderzoek zijn de bevindingen vertaald naar een onderzoeksagenda. Hiervoor zijn de kenmerken van stelsels en mogelijke veranderingen van stelsels geselecteerd die vragen oproepen voor het Nederlandse onderwijs. Voor deze selectie is onder andere gekeken naar de omvang van de effecten en de hardheid van de evidentie. Belangrijk daarbij is ook de toepasbaarheid in Nederland, dat wil zeggen in hoeverre is een stelselverandering te realiseren in Nederland.

3 Stelselkenmerken

In dit hoofdstuk wordt per thema de evidentie op een rij gezet. Omwille van de leesbaarheid begint elke paragraaf met een korte samenvatting van de empirische bevindingen en belangrijkste conclusies. Vervolgens worden de belangrijkste studies besproken. Daarbij wordt ingegaan op de opzet / het design van de studie en worden de belangrijkste uitkomsten gemeld ten aanzien van drie publieke belangen: kwaliteit, toegankelijkheid en doelmatigheid.

3.1 Concurrentie, vrije school keuze en segregatie

In het Nederlandse onderwijs hebben leerlingen/ouders de mogelijkheid om zelf een school te kiezen. Landen zoals de Verenigde Staten of het Verenigd Koninkrijk kennen restricties bij de keuze van de school. Zo wordt in de Verenigde Staten nog heftig gedebatteerd over de introductie van vrije schoolkeuze. Vanuit internationaal perspectief kan het Nederlandse stelsel worden gekarakteriseerd als een stelsel met vrije schoolkeuze. De stelselvragen die in Nederland spelen gaan vooral over het bevorderen dan wel verminderen van de schoolkeuze. De mogelijkheden voor schoolkeuze kunnen worden bevorderd door het opheffen van restricties op het toetreden van scholen of het aanbieden van opleidingen. De vrije schoolkeuze wordt verminderd door het introduceren van restricties, zoals het postcodebeleid in het Amsterdamse basisonderwijs.

Vrije schoolkeuze kan leiden tot concurrentie tussen scholen en ook tot segregatie van bepaalde groepen leerlingen. Concurrentie kan scholen stimuleren de kwaliteit van het onderwijs te verbeteren. De conclusie uit het empirisch onderzoek lijkt te zijn dat concurrentie bijdraagt aan de prestaties van leerlingen. Een vrije schoolkeuze biedt leerlingen en ouders de mogelijkheid om de school te kiezen die het best aansluit bij hun wensen en behoeften. Een betere 'match' tussen leerling en school kan de prestaties van leerlingen verbeteren. Het empirisch onderzoek laat zien dat de prestaties van leerlingen met een lagere sociaaleconomische achtergrond toenemen als zij een kwalitatief betere school bezoeken. Ook blijkt dat landen met meer schoolkeuze betere scores op internationale toetsen. De mogelijke keerzijde van een vrije schoolkeuze is dat hierdoor meer ongelijkheid kan ontstaan, bijvoorbeeld als leerlingen met een lagere sociaaleconomische achtergrond vaker kiezen voor scholen met een lagere kwaliteit. Recent onderzoek voor de Verenigde Staten bevestigt dit: gezinnen met lagere inkomens kiezen vaker voor scholen met een lagere kwaliteit, gemeten in testcores. Het aanbieden van transparante informatie over schoolkwaliteit blijkt invloed te hebben op de schoolkeuze in deze gezinnen. Bovendien blijken de overstappers betere resultaten te halen op scholen met een hogere kwaliteit. Vrije schoolkeuze kan ook leiden tot segregatie. Concentratie van leerlingen met een lage sociaaleconomische achtergrond op bepaalde scholen kan nadelig zijn vanwege zogenaemde 'peereffecten'. Het empirisch

onderzoek levert bewijs voor het optreden van peereffecten, maar deze effecten lijken klein te zijn.

Wat betekent deze empirische evidentie voor het Nederlandse stelsel? Hoewel niet alle resultaten eenduidig zijn, lijkt het algemene beeld uit de literatuur te zijn dat concurrentie en vrije schoolkeuze leiden tot gemiddeld betere prestaties, en derhalve positief bijdragen aan *de kwaliteit van het onderwijs als publiek belang*. Daar staat echter tegenover dat de ongelijkheid in prestaties kan toenemen door vrije schoolkeuze. Als vrije schoolkeuze ertoe leidt dat de sociaaleconomische achtergrond van een leerling samenhangt met de kwaliteit van de school kan gesproken worden van een vermindering van *de toegankelijkheid van het onderwijs*. Het bevorderen van concurrentie en keuzemogelijkheden, bijvoorbeeld door de toetreding van scholen te bevorderen, kan leiden tot betere prestaties. Dit kan echter ook meer ongelijkheid in de hand werken. Hier lijkt derhalve sprake van een uitruil tussen kwaliteit en ongelijkheid. De ongelijkheid kan worden verminderd door het beschikbaar stellen van eenvoudig toegankelijke informatie. Het verschil in keuzegedrag naar sociaaleconomische achtergrond in relatie tot de beschikbaarheid van transparante informatie over kwaliteit van scholen lijkt belangrijk voor het Nederlandse onderwijsstelsel en roept de volgende vragen op:

- Kiezen ouders met een lagere sociaaleconomische achtergrond in Nederland vaker voor scholen met een lagere kwaliteit?
- In hoeverre hangt dit samen met de transparantie en toegankelijkheid van informatie over de kwaliteit van scholen in Nederland?
- In hoeverre leiden verschillen in schoolkeuze naar sociaaleconomische achtergrond tot verschillen in prestatieprikkels voor scholen?
- Zijn scholen met een goede kwaliteit terughoudend in het verspreiden van informatie over hun kwaliteit, om daarmee te voorkomen dat leerlingen van slecht geïnformeerde ouders voor hun school kiezen?

Concurrentie tussen scholen

Als leerlingen en ouders kunnen kiezen tussen scholen, en daardoor 'met hun voeten kunnen stemmen', kan dit scholen stimuleren de kwaliteit van het onderwijs te verbeteren. Onderzoek naar de invloed van concurrentie tussen scholen op leerprestaties is niet eenvoudig vanwege selectieproblemen (leerlingen/ouders kiezen scholen) en omgekeerde causaliteit (goed presterende scholen kunnen groter worden ten koste van slecht presterende scholen waardoor de concurrentie afneemt).

Belfield en Levin (2003) vatten de resultaten van 41 empirische studies over concurrentie in de VS samen. De meeste studies vinden positieve effecten van concurrentie op leerprestaties maar de grootte van de effecten is bescheiden. Wellicht de bekendste studie naar de invloed van concurrentie is gedaan door Caroline Hoxby (2000). Zij maakt gebruik van variatie in concurrentie die voortkomt uit natuurlijke barrières, zoals grote rivieren. Grote rivieren kunnen

de concurrentie tussen scholen (districten) beperken omdat scholen (districten) aan de overzijde van de rivier moeilijk toegankelijk zijn vanwege grotere reiskosten. Tussen situaties met de minste en de meeste keuze vindt zij verschillen in prestaties tussen 0,25 en 0,5 standaarddeviatie. Deze bevindingen zijn echter bekritiseerd in een replicatie door Rothstein (2007). Het belangrijkste kritiekpunt is dat het meten van rivieren en grote stromen weinig precies is. Analyses met alternatieve metingen van rivieren geven veel kleinere effecten van concurrentie. Een studie van Gibbons et al. (2008) maakt gebruik van verschillen in concurrentie tussen gebieden die voortkomt uit de geografische configuratie van scholen. In sommige gebieden liggen scholen allemaal dicht bij het centrum terwijl in andere gebieden scholen meer verspreid zijn en soms aan de rand van de gebieden liggen. Scholen met grotere afstanden zullen minder sterk concurreren. Zij vinden geen effecten van concurrentie op leerprestaties. Noailly et al (2009) passen deze methode toe op het Nederlandse basisonderwijs. Zij vinden kleine positieve effecten van concurrentie op leerprestaties.

De concurrentie tussen scholen kan ook toenemen door een uitbreiding van het aanbod van private scholen of door een verbetering van de toegankelijkheid van private scholen. Holmes et al. (2005) analyseren het effect van de introductie van zogenoemde 'charter schools' op de prestaties in publieke scholen in North Carolina. Zij vinden dat de prestaties in publieke scholen toenemen met 0,25 standaarddeviatie. Hsieh en Urquiola (2003) onderzochten de effecten van de introductie van vrijheid van schoolkeuze in Chili in 1981. Met deze hervorming kwamen publieke subsidies beschikbaar voor private scholen, vergelijkbaar met het huidige Nederlandse stelsel. Zij vinden geen effecten van deze hervorming op de gemiddelde leerprestaties. Wel vinden zij dat de beste leerlingen overstappen van publieke scholen naar private scholen.

Bovenstaande studies lijken er op te wijzen dat concurrentie positief bijdraagt aan leerprestaties. De effecten lijken echter bescheiden te zijn.

Vrije schoolkeuze

Vrije schoolkeuze biedt ouders en leerlingen de mogelijkheid een school te kiezen die het best past bij hun specifieke wensen en behoeften. Een betere 'match' tussen leerling en school kan leiden tot betere prestaties.

Studies die gebruik maken van verschillen tussen landen

Het aanbod van private scholen in een land kan gezien worden als een indicator van vrije schoolkeuze. De invloed van het aanbod van private scholen op de prestaties van leerlingen is in enkele recente studies onderzocht (Woessmann (2001, 2003, 2008, 2009)). De prestaties van leerlingen uit verschillende landen zijn gemeten op internationale toetsen, zoals TIMSS en PISA. Voor zowel TIMMS, PISA 2000 als PISA 2003 wordt gevonden dat landen met een groter aandeel privaat opererende scholen hogere scores halen. Tegelijkertijd blijkt een groter aandeel publieke financiering positief samen te hangen met de testcores. Anders gezegd, landen die een hoog aandeel private scholen combineren met een hoog aandeel publieke

financiering halen de hoogste testcores. Deze resultaten dienen echter met enige voorzichtigheid te worden geïnterpreteerd omdat er vele niet geobserveerde verschillen kunnen bestaan tussen landen, die mogelijk de gevonden resultaten kunnen vertekenen.

Studies die gebruik maken van verschillen binnen landen

In de Verenigde Staten zijn enkele experimenten uitgevoerd waarin de schoolkeuze werd vergroot door het verstrekken van vouchers. Peterson et al. (2003) analyseren de resultaten van experimenten waarin leerlingen vouchers voor deelname aan private scholen wordt aangeboden. Zij vinden dat de prestaties van Afro-Amerikaanse leerlingen verbeteren als zij overstappen van publieke naar private scholen. Een replicatie van Krueger and Zhu (2004) stelt deze resultaten echter ter discussie. Ander bewijs dat meer schoolkeuze leidt tot betere prestaties kan worden ontleend aan een recente studie naar het desegregatie beleid in de Verenigde Staten. Na de uitspraak van het Amerikaanse Hooggerechtshof uit 1954 dat separate scholen voor zwarte en blanke leerlingen inherent ongelijk zijn, kregen Afro-Amerikaanse leerlingen een ruimere schoolkeuze. In veel staten werden zwarte Amerikaanse leerlingen vanuit de binnensteden vervoerd naar scholen in de voorsteden (busing). In een studie van Guryan (2004) is het effect onderzocht van dit desegregatiebeleid op het voortijdig schoolverlaten. In dit onderzoek is gebruik gemaakt van verschillen in de timing van het desegregatiebeleid tussen Amerikaanse staten. De desegregatieplannen uit de jaren zeventig hebben geleid tot een daling van het voortijdig schoolverlaten van zwarte leerlingen van 2 tot 3 procent in de jaren zeventig en tachtig. De desegregatie heeft niet geleid tot een toename van voortijdig schoolverlaten voor blanke leerlingen.

Ook is veel onderzoek gedaan naar verschillen in prestaties op publieke en private scholen. De kernvraag daarbij is of een leerling beter presteert als hij/zij een private school bezoekt dan als hij/zij een publieke school bezoekt. Het probleem bij deze studies is dat leerlingen niet door toeval worden toegewezen aan publieke of private scholen. Verschillende studies hebben getracht om rekening te houden met selectie van leerlingen in private scholen. Neal (1997) vindt dat de keuze voor katholieke scholen leidt tot betere prestaties. Levin (2002) vindt vergelijkbare resultaten voor Nederland. Cullen et al. (2006, 2007) maken gebruik van het systeem van loting dat gebruikt wordt bij de toelating tot 'high schools' in Chicago. Zij vinden geen effect van het volgen van onderwijs op scholen met gemiddeld betere prestaties op toetsresultaten, maar wel een verlagend effect op andere uitkomsten zoals disciplinaire maatregelen en arrestaties.

Studies waarin de prestaties van leerlingen op internationale toetsen in verschillende landen worden geanalyseerd laten zien dat leerlingen beter presteren als hun school privaat bestuurd wordt. Het verschil in prestaties tussen publieke en privaat bestuurde scholen bedraagt gemiddeld 16 tot 20 punten op de PISA-score in drie verschillende testen van PISA 2000 (Fuchs&Woessmann, 2007). Hetzelfde geldt voor PISA 2003 (Woessman et al, 2007). De resultaten blijken echter niet uniform voor alle landen, en af te hangen van de wijze van

financiering van scholen. In landen waarin private scholen afhankelijk zijn van overheidsfinanciering in plaats van private financiering blijken private scholen beter te presteren dan publieke scholen (Corten en Dronkers, 2006; Dronkers en Robert, 2008). Woessmann et al. (2007) vinden ook dat leerlingen in private scholen hoger scoren op niet-cognitieve uitkomsten als 'storend gedrag' en 'op tijd komen'. De resultaten van deze landenstudies kunnen evenwel vertekend zijn door niet geobserveerde selectie van leerlingen in private scholen.

Schoolkeuze, heterogene voorkeuren en informatie

Vrije schoolkeuze vergroot de mogelijkheden om een school kiezen. Dit kan leiden tot meer ongelijkheid als gezinnen met een lagere sociaaleconomische achtergrond vaker kiezen voor scholen met een lagere kwaliteit. De recente plannen in de gemeente Nijmegen, om via een centraal meldpunt te komen tot een andere spreiding van leerlingen over scholen, lijken ook ingegeven door bezorgdheid over de uitkomsten van de vrije schoolkeuze.

Recent onderzoek uit de Verenigde Staten laat zien dat ouders met een lagere sociaaleconomische achtergrond vaker scholen kiezen met een lagere kwaliteit, gemeten met testcores (Hastings & Weinstein, 2008). Zij onderzoeken de invloed van eenvoudig toegankelijke informatie over prestaties van scholen op het keuzegedrag van ouders met een lager of midden inkomen. Ze maken gebruik van een natuurlijk experiment en een gecontroleerd experiment in het Charlotte-Meckleburg schooldistrict. Het natuurlijk experiment maakt gebruik van de situatie dat leerlingen (ouders) in slecht beoordeelde scholen in het Charlotte-Meckleburg schooldistrict hiervan direct op de hoogte werden gesteld. Daarnaast ontvingen zij eenvoudig toegankelijke informatie over de prestaties van andere scholen in de buurt. Ongeveer 15 procent van deze leerlingen stapte over naar een beter scorende school, vooral leerlingen die dicht bij een dergelijk alternatief woonden. In het gecontroleerde experiment werd vervolgens nagegaan of het beschikbaar stellen van eenvoudig toegankelijke informatie vergelijkbare effecten had als de eigen school geen slechte beoordeling had ontvangen. Ook hier werd gevonden dat de groep die informatie ontving vaker koos voor de beter scorende school. De afstand tot de nieuwe school speelde daarbij ook een rol. Vervolgens werd gekeken naar de invloed van de beter scorende school op de prestaties van de leerlingen die waren overgestapt na het krijgen van informatie. Gevonden werd dat deze leerlingen significant betere prestaties haalden. De auteurs concluderen dat schoolkeuze de prestaties van leerlingen met een lagere sociaaleconomische achtergrond kan verbeteren als ouders gemakkelijk toegang hebben tot informatie over prestaties van scholen en als er goede alternatieve scholen in de buurt zijn.

Onderzoek naar schoolkeuzegedrag in Engeland en Frankrijk geeft vergelijkbare resultaten als in de Verenigde Staten. Prestatie-indicatoren van scholen blijken nauwelijks van invloed te zijn op de schoolkeuze. Karsten et al. (2001) concluderen op basis van een aantal internationale

studies dat ouders over het algemeen weinig aandacht besteden aan prestatie-indicatoren.

Volgens Janssens en Visscher (2004) spelen daarbij de volgende factoren een rol:

- Veel ouders hebben geen toegang tot deze informatie of weten niet van het bestaan;
- Het gebruik van de beschikbare informatie vergt inzicht en ervaring, in een aantal gevallen is de informatie ook lastig te vertalen naar een individuele situatie.

Het feit dat gezinnen met lagere inkomens, vergeleken met gezinnen met hogere inkomens, bij de keuze van een school minder gewicht toekennen aan de onderwijsuitkomsten heeft enkele gevolgen. Allereerst is dit nadelig voor de prestaties van leerlingen uit deze groepen. In de tweede plaats vermindert dit de (competitieve) druk op scholen om goede leerprestaties te genereren. De druk op scholen om goede prestaties te leveren is het grootst bij scholen met veel kinderen uit hoge inkomensgroepen en het laagst bij scholen met veel kinderen uit lage inkomensgroepen. Hierdoor kan de ongelijkheid in leerprestaties naar sociaaleconomische achtergrond groter worden. Een derde mogelijk gevolg is dat scholen met een goede kwaliteit terughoudend kunnen zijn met het verspreiden van informatie over hun kwaliteit. Daarmee voorkomen zij dat leerlingen van slecht geïnformeerde ouders voor hun school kiezen. Aannemelijk is dat het opleiden van deze leerlingen grotere inspanningen van de school vergt dan het opleiden van leerlingen met goed geïnformeerde ouders. De auteurs concluderen dat schoolkeuze, zonder additionele prikkels, juist kan leiden tot een grotere stratificatie tussen hoge en lage inkomens groepen, in plaats van meer concurrentie waarvan iedereen profiteert.

Segregatie

Vrije schoolkeuze biedt de mogelijkheid om scholen te kiezen ook buiten de eigen wijk of buurt. Dit kan leiden tot segregatie van leerlingen, bijvoorbeeld naar sociaaleconomische achtergrond of naar etniciteit (witte/zwarte scholen). In het beleid bestaat grote zorg dat concentratie van leerlingen met een lagere sociaaleconomische achtergrond of niet-Nederlandse etniciteit leidt tot zwakke onderwijsuitkomsten. Anders gezegd, de samenstelling van de klas (peers) kan van invloed zijn op de prestaties van een leerling.

Samenstelling van de klas (peereffecten)

Het vaststellen van peereffecten is methodologisch gezien lastig vanwege allerlei selectieprocessen van ouders (leerlingen) en scholen. Door deze selectie is de samenstelling van een klas niet toevallig, waardoor een regressie van onderwijsuitkomsten op klassensamenstelling niet een schatting van het oorzakelijk effect hoeft te geven. Daarnaast is de vraag hoe de invloed van de peers verloopt: wie beïnvloedt wie?

Onderzoek naar de invloed van peers beperkt zich niet tot het onderwijs maar is veel breder en richt zich bijvoorbeeld ook op de invloed van de buurt of de invloed van collega's op het werk. Eén van de meest opvallende studies op dit terrein is het zogenoemde Moving to

Opportunity (MTO) project. In dit project zijn woonvouchers verstrekt aan arme gezinnen in Boston in de Verenigde Staten. Met deze vouchers werden gezinnen in staat gesteld te verhuizen naar veiliger wijken. Door de vouchers random toe te wijzen kon men de effecten van de buurt vaststellen op een groot aantal uitkomsten (Kling, et al. 2007). Vijf tot zeven jaar na de toekenning van de vouchers werden positieve effecten gevonden op de onderwijsdeelname en prestaties van meisjes, daar stonden echter negatieve onderwijsuitkomsten voor jongens tegenover.

In de recente onderwijsliteratuur zijn peereffecten onderzocht door te zoeken naar situaties waarin sprake is van toevallige variatie in de klassensamenstelling (natuurlijke experimenten). Zo is bijvoorbeeld gebruik gemaakt van variatie door veranderingen tussen cohorten van de populatie (Hoxby, 2000, Ammermueller & Pischke, 2006, Lavy & Schlosser, 2007), variatie door de plotselinge instroom van immigranten (Gould, Lavy, Paserman, 2004), variatie door random toewijzing van kamergenoten (Sacerdote, 2000), variatie door vroege selectie (tracking) van leerlingen (Lefgren, 2004) en variatie door spreiding van leerlingen (busing) (Angrist & Lang, 2004). Driessen (2007) geeft een compleet overzicht van deze recente literatuur. Hoewel de resultaten niet eenduidig zijn, lijkt de belangrijkste conclusie te zijn dat peereffecten positief samenhangen met leerprestaties (betere peers leiden tot betere prestaties) maar dat de effecten klein zijn. Veranderingen in klassensamenstelling kunnen daarmee leiden tot een verbetering of verslechtering van prestaties van bepaalde groepen. Voor het beleid levert dit een dilemma. Herschikking van leerlingen kan dan winst opleveren voor bepaalde groepen leerlingen maar daar staan slechtere prestaties van andere leerlingen tegenover. Als peereffecten niet-lineair zouden zijn is het wel mogelijk om via herschikking van leerlingen de totale resultaten te verbeteren zonder dat bepaalde groepen slechter gaan presteren. Het aantonen van niet-lineaire peereffecten is echter gecompliceerd en vereist veel gegevens. Recent zijn twee studies uitgevoerd naar niet-lineaire effecten. Ding et al. (2007) vinden, voor een steekproef van leerlingen uit het voorgezet onderwijs in China, dat goede leerlingen meer dan matige leerlingen profiteren van goede medeleerlingen. Hoxby en Weingarh-Salyer (2005) analyseren herplaatsingen van leerlingen in een groot schooldistrict in North-Carolina. Zij vinden dat leerlingen het sterkst beïnvloed worden door leerlingen die in de beginsituatie sterk op hen lijken. Ook vinden zij dat beperkte heterogeniteit in een klas niet van invloed is op de prestaties maar dat veel heterogeniteit een negatief effect heeft op de prestaties van alle leerlingen in de klas.

De literatuur over peereffecten gaat vrijwel geheel over de effecten op cognitieve leerprestaties. Peereffecten kunnen echter ook een rol spelen bij de vorming van normen, waarden, motivaties en ambities. Neidell en Waldfogel (2008) onderzoeken peereffecten in de voor- en vroegschoolse educatie. Zij vinden wel peereffecten op cognitieve uitkomsten als rekenen en taal, maar geen peereffecten op gedrags- en sociale uitkomsten.

De conclusie bij de huidige stand van de literatuur lijkt te zijn dat peereffecten een rol spelen maar dat de effecten klein zijn. De cognitieve prestaties van een leerling verbeteren in

klassen met gemiddeld betere medeleerlingen. Twee recente studies vinden bewijs voor niet-lineaire peereffecten: leerlingen worden vooral beïnvloed door vergelijkbare medeleerlingen. Welke implicaties deze recente studies hebben voor het Nederlandse onderwijsstelsel is niet duidelijk.

3.2 Autonomie en verantwoording

Vanuit internationaal perspectief kennen Nederlandse scholen een grote mate van autonomie. Autonomie is direct verbonden met verantwoording en toezicht, bijvoorbeeld door het afnemen van centrale examens, het beoordelen van prestaties van scholen of door de Onderwijsinspectie. Empirisch onderzoek laat zien dat de prestaties van leerlingen beter zijn in landen waarin scholen autonomie hebben over het onderwijsproces en over personele beslissingen. Centrale examens blijken een sterk positief effect te hebben op de prestaties van leerlingen. Dit roept de vraag op of het gebruik van schoolexamens in het Nederlandse voortgezet onderwijs, de positieve effecten van centrale examens vermindert. In de VS worden sinds enkele jaren verschillende zogenoemde ‘school accountability’ systemen gebruikt. Dit zijn systemen waarin de prestaties van scholen worden beoordeeld volgens van tevoren vastgelegde standaarden en waarin consequenties worden verbonden aan het niet halen van de normen. Evaluaties van deze systemen wijzen op positieve effecten maar laten ook zien dat strategisch gedrag van scholen veel voorkomt. Een goed design van dit soort systemen is derhalve van groot belang. Welke aanpak voor ‘falende of slecht presterende scholen’ het beste werkt is niet bekend.

Wat betekent deze evidentie voor het Nederlandse stelsel? Het empirisch onderzoek laat zien dat autonomie van scholen en verantwoording door centrale examens of ‘school accountability systemen’ positief bijdragen aan *de kwaliteit van het onderwijs*. Strategisch gedrag van scholen is echter een ‘fact of life’, en kan nadelig zijn voor *de toegankelijkheid* van het onderwijs als scholen bepaalde leerlingen weren. Voor de onderzoeksagenda lijken vooral vragen aan de orde over:

1. Het ontwerp van goede beoordelingssystemen van scholen;
2. De aanpak van falende of slecht presterende scholen.

3.2.1 Autonome scholen

Variatie in autonomie van scholen wordt vooral gevonden tussen landen. In een aantal recente studies is gekeken naar de invloed van de autonomie van scholen op de prestaties van leerlingen uit verschillende landen op internationale toetsen. Hiervoor worden regressieanalyses uitgevoerd van de toetscores op kenmerken van onderwijsstelsels van landen met zoveel mogelijk controlevariabelen voor verschillen tussen leerlingen. Het algemene beeld van dit onderzoek is dat leerlingen in landen met autonomie over het onderwijsproces en over personele beslissingen hogere scores halen (Woessmann, 2003a, 2007, Fuchs & Woessmann,

2007). De beslissingen die scholen kunnen nemen hebben betrekking op het aankopen van voorraden, de verdeling van budgetten binnen scholen, het aannemen en belonen van leraren (bij gegeven budget) en het kiezen van tekstboeken en instructiemethoden. De positieve effecten van autonomie op de prestaties worden zowel in het voortgezet onderwijs als in het basisonderwijs gevonden (Fuchs & Woessman, 2005). Ook blijken leerlingen beter te presteren als hun docenten de mogelijkheid hebben om zelf de lesmethoden te selecteren en docenten beoordeeld worden op basis van hun prestaties (Woessmann, 2003, Fuchs & Woessmann, 2007). De invloed van de autonomie van scholen hangt ook af van het gebruik van centrale eindexamens. Autonomie van scholen heeft een sterker effect op de prestaties van leerlingen in landen met centrale eindexamens (Woessmann, 2003, 2005, Fuchs & Woessmann, 2007, Woessman et al. 2007). De effecten van centrale eindexamens worden in de volgende paragraaf verder besproken. Zoals hiervoor als is opgemerkt, moeten de resultaten van deze vergelijkingen tussen landen met voorzichtigheid worden geïnterpreteerd vanwege mogelijke niet geobserveerde verschillen tussen landen.

Binnen landen bestaat soms variatie in autonomie tussen publieke en private scholen. Voor privaat geleide scholen geldt vaak een grotere autonomie. In de vorige paragraaf zijn studies besproken naar de verschillen in effectiviteit van publieke en private scholen. Deze studies suggereren dat leerlingen betere prestaties halen als ze naar private scholen gaan.

3.2.2 Verantwoording

Voor het afleggen van verantwoording is de meting van prestaties van groot belang. Centrale eindexamens kunnen voorzien in een dergelijke objectieve meting. In de VS worden sinds enkele jaren verschillende zogenoemde 'school accountability' systemen gebruikt. In Nederland speelt de Inspectie van het Onderwijs een centrale rol bij het beoordelen van de prestaties van scholen.

Centrale eindexamens

Centrale examens kunnen zorgen voor transparantie en verantwoording. Dit geeft het schoolpersoneel prikkels om zich te concentreren op het verbeteren van prestaties van leerlingen. Zonder de juiste prikkels zouden leraren mogelijk een voorkeur kunnen hebben voor de meest gemakkelijke lesmethoden in plaats van de meest veelbelovende. Centrale examens geven ook prikkels voor leerlingen om te leren. Wanneer absolute normen ontbreken en er alleen sprake is van relatieve prestatiemeting kunnen leerlingen baat hebben bij een gezamenlijk afspraak om niet te studeren.

Empirische studies bevestigen deze theoretische verwachtingen. Het onderzoek betreft vooral studies waarin de prestaties van leerlingen in landen met een centraal examen worden vergeleken met de prestaties van leerlingen in landen zonder een centraal examen.

Bishop (1997) vergelijkt de internationale TIMMS en IAEP (International Assessment of Educational Progress) toetsscores tussen landen met en zonder centraal examen.⁴

Regressieanalyses op landenniveau laten zien dat TIMMS toetsscores op wiskunde en natuurkunde (science) significant hoger zijn in landen met een centraal examen. Op de IAEP toetsscores wordt een groot positief statistisch significant effect gevonden op wiskunde en een wat minder groot, niet significant effect op natuurkunde (science). Naast deze internationale analyses presenteert Bishop (1997) ook een analyse waarin hij gebruik maakt van variatie tussen verschillende provincies in Canada. Canada heeft een gemengd systeem, waarin sommige provincies centrale examens kennen en sommige niet. Regressies op schoolniveau laten zien dat scholen met een centraal examen significant beter presteren op zowel wiskunde als science. Deze resultaten suggereren dat onderwijsprestaties aanzienlijk kunnen worden verbeterd door de disciplinerende werking van centrale examens.

Woessman (2003) analyseert individuele leerlingprestaties op de TIMMS. Individuele toetsscores worden geregresseerd op persoonlijke achtergrondkenmerken, inzet van middelen, en instituties, waaronder de aanwezigheid van centraal examens. Hij vindt statistisch significante positieve effecten van centrale examens op de toetsscores op wiskunde en science. De grootte van het gevonden effect is respectievelijk 0.16 en 0.10 standaarddeviatie. Een vergelijkbare analyse is uitgevoerd op de PISA-toetsen (Fuchs en Woessman, 2007). In deze studie worden positieve effecten van centrale examens gevonden op alle toetsen. Jürges et al. (2005) maken gebruik van regionale variatie in schoolwetgeving in het Duitse voortgezet onderwijs om het effect van centrale examens op onderwijsprestatie te identificeren.⁵ Hiervoor kijken ze naar de effecten op de TIMMS toetsscores op wiskunde en science. In alle Duitse provincies zijn centrale examens voor wiskunde, maar in de meeste provincies is geen centraal examen voor science. Zij vinden significant positieve effecten van centrale examens op toetsscores. Afhankelijk van de precieze steekproefdefinitie en specificatie varieert de grootte van het effect van een vijfde tot tweederde van een grade level equivalent.

Deze empirische studies laten een eenduidig beeld zien wat betreft de effecten van centrale examens op de kwaliteit van het onderwijs. Centrale examens kunnen onderwijsprestaties aanzienlijk verbeteren en lijken een relatief kosteneffectieve maatregel. Woessman (2003) laat zien dat een centraal examen veel kosteneffectiever is dan klassenverkleining.

In het Nederlandse voortgezet onderwijs wordt het centraal examen gebruikt in combinatie met schoolexamens. Dronkers en De Lange (2006) hebben discrepanties in de becijfering van deze twee soorten examens geanalyseerd tussen 1998 en 2005 en concluderen dat de prestaties op deze beide typen examens meer uiteen zijn gaan lopen.

⁴ De internationale TIMMS data bevatten testresultaten van 13-jarige leerlingen op wiskunde en science. De internationale standaardfout is 100 punten en een klasniveau is ongeveer gelijk aan 40 punten.

⁵ Jürges et al. (2005) maken evenals Bishop (1997) gebruik van regionale variatie. Zij stellen zich de vraag of de analyse van Bishop daadwerkelijk een causaal effect oppikt. Zo is het in de analyse op Canadese scholen bijvoorbeeld denkbaar dat zowel een centraal examen als hogere toetsscores het gevolg zijn van de grotere waarde die in bepaalde provincies gehecht wordt aan onderwijsprestaties.

Het gebruik van 'school accountability' systemen

De invloed van systemen voor het beoordelen van de prestaties van scholen zijn vooral onderzocht in de Verenigde Staten maar er is ook enige evidentie uit landenstudies.

In een vergelijking van de prestaties van leerlingen op internationale toetsen vinden Woessman et al. (2009) dat de prestaties hoger zijn in landen waarin beoordelingen van individuele scholen worden vergeleken met prestaties in het district of op landelijk niveau. Openbaarheid van prestatiegegevens blijkt een positief effect te hebben op de prestaties van scholen. Twee studies (Hanushek and Raymond (2005) en Jacob (2005)) wijzen er op dat de invoering van het 'school accountability' systeem in the VS, het 'No Child Left Behind Act', een positief effect heeft op leerprestaties. Hierbij maken ze gebruik van toevallige variatie tussen staten in het jaar waarin ze verschillende onderdelen van dit accountabilitybeleid hebben geïntroduceerd. Ook hanteren ze aanvullende statistische methoden om beleidseffecten van andere effecten te onderscheiden. Beide studies tonen aan dat de introductie van het accountabilitybeleid heeft geresulteerd in een toename in toetsscores. Interessante verschillen tussen subgroepen kwamen ook aan het licht; 'hispanic and white students gain much more than black students' (Hanushek & Raymond, 2005). Ze concludeerden hierbij dat het accountability systeem tot betere leerresultaten leidt, maar tegelijkertijd de verschillen tussen groepen studenten (de 'black-white gap') vergroot. Beide studies onderzochten ook welke causale mechanismen verantwoordelijk waren voor de effectiviteit van het accountability-systeem. Hanushek and Raymond (2005) onderzochten of de introductie van de publieke 'school report cards' hiervoor verantwoordelijk was, maar vonden geen afzonderlijk significant effect op leerprestaties. Een uitgebreidere bespreking van deze en andere studies naar effecten van accountability-systemen is te vinden in De Wolf en Janssens (2007).

School-accountability systemen kunnen ook strategisch gedrag in de hand werken. Jacob (2002) onderzoekt het accountability beleid dat in 1996-1997 in publieke scholen in Chicago is ingevoerd. Scholen werden onder toezicht geplaatst als minder dan 15 procent van de leerlingen de nationale standaarden haalde. Als zich vervolgens geen verbetering voordeed werden docenten en schoolleiders herplaatst of ontslagen. Hij vindt een aanzienlijke verbetering van testcores in wiskunde en lezen. Echter, er lijkt sprake van 'teaching-to-the test'. Er was geen verbetering op toetsen die geen deel uitmaakten van het accountability beleid. Daarnaast werden meer leerlingen verwezen naar trajecten die niet getest werden, bijvoorbeeld naar het speciaal onderwijs. Figlio & Getzler (2002) onderzochten de effecten van een nieuw toetsing systeem dat in 1996/1997 werd geïntroduceerd in Florida. Zij vinden dat de introductie van het nieuwe accountability systeem geleid heeft tot het vaker classificeren van leerlingen als gehandicapt, en dat dit zich vooral voordeed in scholen met grote aandelen leerlingen uit lagere inkomensgroepen. Jacob & Levitt (2003) ontdekten dat accountability systemen leidden tot fraude door docenten. Om fraude te ontdekken maakten ze gebruik van een algoritme dat ongebruikelijke antwoordpatronen op testen kon opsporen. Fraude bleek zich jaarlijks voor te doen in 4 tot 5 procent van de basisscholen. Ze tonen ook aan dat dit de scholen zijn waar de

fluctuatie in toetsscores over de jaren heen groter is dan gemiddeld. Figlio (2005) vindt bewijs dat scholen schorsingen strategisch inzetten. Gedurende de afname van toetsen in het kader van accountability systemen zijn laag presterende leerlingen vaker geschorst dan gedurende de rest van het jaar. Dit verschil treedt alleen op voor laag presterende leerlingen in de groepen die getest worden. Burgess et al. (2005) onderzoeken de effecten van accountability beleid in het Verenigd Koninkrijk. Dit beleid zet scholen aan om zich vooral in te spannen voor leerlingen die zich in de marge bevinden van het halen van de standaard. Zij vinden dat de prestaties van leerlingen met een lage geschiktheid slechter waren na de introductie van het nieuwe beleid. In het Nederlandse onderwijs speelt strategisch gedrag van scholen mogelijk een rol bij verschillen in deelname aan de CITO-toets en bij verschillen tussen cijfers voor het centraal examen en cijfers voor het schoolexamen (Dronkers en De Lange, 2006).

Inspectie toezicht

De effecten van toezicht door de Inspectie zijn vooral onderzocht in Engeland en recent ook in Nederland. Het onderzoek kijkt naar het effect van een belangrijk instrument van het toezicht, namelijk het effect van schoolbezoeken. Slechts één studie tracht de invloed van het bestaan van een Onderwijs Inspectie als onderdeel van een onderwijsstelsel te onderzoeken. Ladd (2009) vergelijkt het accountability-systeem in de Verenigde Staten met het inspectiesysteem in Nederland en Nieuw Zeeland. Ze vindt dat de prestaties in deze laatste twee (inspectie)landen hoger zijn dan in de Verenigde Staten. Het blijft evenwel de vraag in hoeverre deze verschillen het gevolg zijn van het bestaan van de Onderwijsinspectie. Enkele studies hebben het effect van schoolbezoeken door inspecteurs op de leerlingprestaties onderzocht. Drie studies onderzoeken de invloed van bezoeken door de Engelse Onderwijsinspectie op de resultaten van scholen in het voortgezet onderwijs (Cullingford et al. 1999, Wilcox and Gray 1996, and Shaw et al. 2003). Alle drie studies vinden een negatief effect van deze bezoeken. Deze studies kunnen echter vertekend zijn door selectie van scholen door de inspecteurs, zij hebben mogelijk een voorkeur voor het bezoeken van zwakkere scholen. Een studie van Rosenthal (2004) maakt gebruik van panelgegevens van Engelse scholen in het voortgezet onderwijs. Ook deze studie vindt een negatief effect van inspectiebezoeken. Luginbuhl et al (2009) onderzoeken de invloed van bezoeken van onderwijsinspecteurs in het Nederlandse basisonderwijs. Zij maken daarbij gebruik van een natuurlijk experiment. In het kader van het jaarlijkse onderwijsverslag trekt de Inspectie van het onderwijs een random steekproef van scholen die worden bezocht. De invloed van deze random bezoeken op de prestaties van scholen op de CITO-eindtoets is onderzocht. Daarnaast zijn fixed-effect modellen geschat voor alle inspectiebezoeken. Voor de random bezoeken worden geen effecten gevonden, voor de fixed-effect schattingen worden kleine positieve effecten van schoolbezoeken gevonden.

Enkele studies richten zich ook op de effecten van toezicht op schoolbeleid en schoolverbetering. Geen van deze studies hanteert een (quasi)experimenteel design (zie voor een overzicht, De Wolf en Janssens (2007) en Ehren (2007)). De meeste studies laten een

positief effect zien. Zo blijkt toezicht op 50 tot 70% van de scholen tot gedragsbeïnvloeding en beleidsveranderingen leidt. De omvang van de gevonden effecten is echter sterk afhankelijk van het onderzoeksdesign.

Aanpak van falende of slecht presterende scholen

Over een effectieve aanpak van falende of slecht presterende scholen bestaat nog weinig harde evidentie. Zowel in Engeland (Matthews & Sammons, 2004) als in Nederland (Janssens en De Wolf, 2008; Inspectie van het Onderwijs, 2007) bestaan studies die aantonen dat zeer zwakke scholen in staat zijn zich te verbeteren. Over wat deze verbetering precies heeft veroorzaakt, is weinig hard bewijs. Deze studies hanteren geen quasi-experimentele onderzoeksdesigns.

De aanpak van falende of slecht presterende scholen verschilt sterk tussen landen. Engeland kent een uitgebreid beleid op dit terrein. Op alle scholen die onderpresteren wordt het toezicht geïntensiveerd. Daarnaast wordt onderscheid gemaakt in scholen waarvan verwacht wordt dat ze verbeteringen zelf kunnen doorvoeren en scholen waar dit niet het geval is. Scholen die zelf geacht worden er bovenop te komen krijgen een zogenaamde 'notice to improve'. Van de gemeente wordt verwacht dat ze de school hierbij ondersteunen. Bij negen van de tien falende scholen blijkt een verbetering op te treden. De Engelse inspectie van het onderwijs (Ofsted) onderscheidt daarnaast een groep falende scholen waarvan de leiding niet de capaciteiten heeft om de nodige verbeteringen door te voeren. Voor deze scholen geldt speciaal beleid. De 'local authority' wordt hier om een verbeterplan gevraagd. Dit laatste kan bijvoorbeeld inhouden dat de gemeente de school extra middelen verstrekt, een nieuw schoolhoofd en/of nieuw personeel aanstelt of succesvolle scholen de falende school laat ondersteunen. Met intensieve en frequente inspecties wordt het verbeterproces gemonitord. Als een school na 2 jaar nog steeds onderpresteert, zijn er drie verschillende opties:

- De school wordt gesloten;
- Er wordt een nieuwe school geopend, met een nieuw hoofd en niet personeel;
- Er wordt een zogenaamde 'Academy' geopend, veelal moderne brede scholen die opgezet worden door sponsors met verschillende achtergronden (colleges, universiteiten, bedrijven, etc.).

De toezichthouder speelt een belangrijke rol in het Engelse model. Dit is minder het geval in de meeste Amerikaanse staten. Daar bepalen leerprestaties sinds de 'No Child Left Behind'-Act of een school wel of niet onderpresteert. Prikkel en sancties zijn veelal financieel van aard. In een aantal gevallen worden ouders ook in de gelegenheid gesteld alternatieve scholen te bezoeken wanneer de school van hun kind onderpresteert. In een beperkt aantal staten is men begonnen met het instellen van inspecties voor scholen die zwak presteren. Over de effectiviteit van de verschillende aanpakken is weinig bekend. Over het algemeen wordt verondersteld dat sterke consequenties effectiever zijn (Ayres & Braithwaite, 1992). Ook de eerste voorzichtige

resultaten van het International Project for the study of Educational Accountability wijst in deze richting. Maar ook andere factoren kunnen een rol spelen. Zo blijkt uit diverse experimenten dat de manier waarop gehandhaafd wordt (onder andere de communicatie, zie Van Erp, 2008) en ook beloning (zie van der Pligt, 2008) effectieve interventiemechanismen zijn.

Intern gebruik van assessmentresultaten / opbrengstgericht werken

Het in scholen en klassen gebruiken van assessment- en toetsresultaten blijkt de kwaliteit van het onderwijs te vergroten. Het gaat hier weliswaar niet om een formeel stelselkenmerk, maar opbrengstgericht werken draagt bij aan een effectief stelsel. De conclusie mogen we trekken op basis van drie belangrijke overzichtsartikelen; Kluger & DeNisi (1996), Black & William (1998) en Hattie en Imperley (2007). De auteurs komen allen op basis van meta-analyses en literatuurstudies tot de bevinding dat feedback en gebruik van assessment- en toetsresultaten op klasniveau leiden tot effectiever onderwijs. Hoe opbrengstgericht werken het beste op stelselniveau gestimuleerd kan worden, is niet bekend.

3.3 Selectie

De selectie in het Nederlandse onderwijs vindt vooral plaats bij de overgangen tussen onderwijstypen. In de eerste plaats gaat het dan om de overgang van primair naar voortgezet onderwijs. Na het basisonderwijs stromen leerlingen door naar verschillende niveaus van voortgezet onderwijs. Andere landen, zoals de Verenigde Staten of de Scandinavische landen, selecteren pas op latere leeftijd. Het Nederlandse onderwijs kent derhalve een vroege selectie van leerlingen. Andere overgangen doen zich voor na het afronden van verschillende vormen van voortgezet onderwijs. De internationale empirische literatuur kijkt vooral naar de invloed van een vroege selectie van leerlingen. Er bestaat geen eenduidig beeld van de invloed van vroege selectie op de kwaliteit van het onderwijs. Wel lijkt het empirisch onderzoek er op te wijzen dat vroege selectie van leerlingen de invloed van de sociaaleconomische achtergrond van leerlingen op hun onderwijsuitkomsten versterkt. Anders gezegd, vroege selectie lijkt te leiden tot meer ongelijkheid in onderwijskansen, en vermindert derhalve *de toegankelijkheid*. Voor Nederland is gevonden dat uitstel van selectie, voor leerlingen met een Mavo-advies, kan leiden tot meer hoger opgeleiden. De empirische literatuur geeft geen antwoord op de vraag in hoeverre latere doorstroommogelijkheden eventuele nadelige effecten van vroege selectie ongedaan kunnen maken.

Voor Nederland lijkt de vraag aan de orde hoe mogelijke nadelige effecten van vroege selectie het best kunnen worden tegengegaan, hetzij door uitstel van selectie van leerlingen, hetzij door het vergroten van latere doorstroommogelijkheden.

3.3.1 Vroege selectie

Op voorhand zijn de effecten van vroege selectie niet direct duidelijk. Er kan sprake zijn van een tradeoff tussen efficiëntie en ongelijkheid. Enerzijds kan vroege selectie leiden tot meer efficiëntie vanwege een grotere homogeniteit van de klassen. In een homogene klas is het gemakkelijker het lesmateriaal af te stemmen op de behoefte van leerlingen. Anderzijds kan vroege selectie nadelig zijn voor relatief zwakkere leerlingen. Dit kan veroorzaakt worden door bijvoorbeeld verschillen in kwaliteit van leraren of peer group effecten. Vroege selectie van leerlingen brengt meer onzekerheid met zich wat betreft de werkelijke mogelijkheden van een kind, waardoor er een groter risico bestaat dat de leerling in een verkeerd type onderwijs terechtkomt.

Verschillende recente studies onderzoeken de effecten van vroege selectie. Ariga en Brunello (2007) maken gebruik van internationale data en bestuderen het effect van de duur van selectie op de prestaties van jongvolwassenen op cognitieve toetsen. Hiertoe verdelen ze het totaal aantal gevolgde jaren onderwijs in 'het aantal jaren op een geselecteerd niveau' en 'het aantal jaren in een klas met gemengd niveau'. Het aantal jaren in een selectief systeem hangt zowel af van exogene variatie in het tijdstip van selectie tussen landen als endogene variatie binnen landen veroorzaakt door verschillen in leeftijd, schooluitval en verschillen in curriculum. Met mogelijke zelfselectie wordt omgegaan door een instrumentele variabele aanpak waarbij het aantal jaren in een selectief systeem wordt geïnstrumenteerd door dummy variabelen voor leerlingen die zijn gestopt met school vanwege financiële beperkingen, familieomstandigheden, of vanwege persoonlijke beperkingen zoals ziekte. De instrumentele variabele schattingen suggereren een positieve bijdrage van het aantal jaren in een selectief systeem aan onderwijsprestaties.

Duflo et al. (2008) vinden aanwijzingen voor een positief effect van selectie op onderwijsprestaties in een gerandomiseerd experiment in Kenia. Hier kregen 121 scholen de mogelijkheid een extra leraar in te huren om een klas in twee groepen te splitsen. In 60 aselect gekozen scholen werden leerlingen random geselecteerd in één van de twee groepen. In de 61 andere scholen werden de leerlingen gerangschikt op basis van eerdere studieresultaten, waarbij de bovenste helft in de ene groep en de onderste helft in de andere groep werd geplaatst. Na 18 maanden bleken leerlingen in de scholen die selecteerden naar vaardigheden gemiddeld 0.14 standaarddeviaties hoger te scoren dan leerlingen in de scholen die random selecteerden. Dit effect bleef ook een jaar later nog overeind. Ze vonden bovendien dat leerlingen op alle niveaus van de verdeling profiteerden van selectie.

Van Elk et al. (2009) onderzoeken het effect van vroege selectie op de deelname aan en het afronden van het hoger onderwijs in Nederland. Hiervoor benutten ze variatie in het tijdstip van selectie. Op sommige scholen worden leerlingen direct geselecteerd in categorale klassen van een bepaald onderwijsniveau, terwijl andere scholen één- of tweejarige gecombineerde brugklassen aanbieden, waarin leerlingen bij elkaar gehouden worden voordat ze naar een specifiek onderwijsniveau gaan. De keuze voor een bepaald onderwijsniveau wordt hiermee

met één of twee jaar uitgesteld. Met mogelijke zelfselectie wordt op twee manieren omgegaan. Allereerst wordt gecontroleerd voor veel individuele achtergrondkenmerken, zoals testcores en sociaaleconomische achtergrond. Daarnaast maken zij gebruik van exogene variatie veroorzaakt door het aanbod van verschillende typen scholen in de regio van een leerling. Beide methoden laten een statistisch significant negatief effect van vroege selectie op de deelname aan en het behalen van het hoger onderwijs zien.

Hanushek en Woessman (2006) benutten variatie in het tijdstip van selectie tussen landen. Ze gebruiken een difference-in-differences methode om het effect van een vroeg selectie regime op gestandaardiseerde testcores te identificeren. Hiertoe vergelijken ze verschillen in prestaties op de internationale PISA toetsscores in het voortgezet onderwijs en toetsscores in het primair onderwijs tussen landen. Een onderwijssysteem wordt gedefinieerd als een vroeg selectie regime als leerlingen geselecteerd worden voordat ze de test in het voortgezet onderwijs maken. Ze vinden dat vroege selectie de ongelijkheid in toetsresultaten vergroot. Vroeg selecteren lijkt ook de gemiddelde prestatie iets te verlagen, maar dit effect is minder duidelijk. Zo wordt een negatief effect op lezen en een positief effect op science gevonden.

Figlio en Page (2002) onderzoeken het effect van selectie op het verschil in toetsresultaten op wiskunde toetsen tussen het 8e en 10e leerjaar voor groepen van leerlingen met een verschillend niveau. Op basis van de scores in het 8e jaar verdelen ze de leerlingen in de drie categorieën: leerlingen in de hoogste, middelste en laagste derden van de verdeling van toetsscores. Ze schatten verschillende regressies voor elk van deze groepen. Hierbij maken ze gebruik van variatie in selectie tussen scholen.⁶ Ze vinden geen significant verschil in elk van deze regressies en interpreteren dit als een aanwijzing dat selectie geen schadelijke gevolgen heeft voor de zwakkere leerlingen.

Invloed vroege selectie op ongelijkheid/toegankelijkheid

De discussie over effecten van selectie op ongelijkheid wordt vaak gerelateerd aan sociaal-economische achtergrond. Brunello en Checchi (2007) gebruiken internationale data voor verschillende cohorten van leerlingen en richten zich op de interactie tussen familiale achtergrond en selectie. Ze vinden enige evidentie dat selectie de rol van sociaaleconomische achtergrond op het uiteindelijke onderwijsniveau en arbeidsmarktuitskomsten vergroot.

Schuetz et al. (2005) vinden ook dat vroege selectie de invloed van sociaaleconomische achtergrond versterkt en de spreiding in leerlingprestaties doet toenemen. Zij volgen een vergelijkbare benadering als Brunello en Checchi (2007) waarin zij zich richten op het effect van de interactie tussen sociaaleconomische achtergrond en een indicator voor vroege selectie

⁶ In de US is strikt genomen sprake van een gemengd schoolsysteem, waarbij er verschillende niveaustromen worden aangeboden binnen scholen. Dit is een mildere vorm van selectie van leerlingen op basis van vaardigheden. Dit verschil in definitie van selectie wordt vaak genegeerd in de literatuur, maar een probleem met studies uit de VS is wel dat er geen eenduidige definitie van selectie bestaat. Figlio en Page (2002) nemen selectie op als een dummy variabele die aangeeft of de directeur van de school aangeeft of er selectie plaatsvindt. Daarnaast voeren zij nog enkele robuustheidsanalyses uit waarin ze alternatieve definities van selectie opnemen.

(waarvoor zij de leeftijd van selectie nemen) op toetscores van leerlingen in meer dan vijftig landen.

Ammermuller (2005) onderzoekt of verschillen in gelijkheid van onderwijskansen tussen landen gerelateerd zijn aan bepaalde kenmerken van onderwijssystemen. Hij gebruikt internationaal vergelijkbare micro data van de PIRLS en PISA studies naar onderwijsprestaties in 14 landen voor een difference-in-differences analyse, waarbij verschillen tussen toetscores in grade 4 (het einde van het primair onderwijs) en grade 9 (rond het einde van de verplichte onderwijs) vergeleken worden tussen landen met verschillende instituties. Hierbij vindt hij dat met name leerlingen met een betere sociale achtergrond profiteren van differentiatie van leerlingen naar vaardigheden. Gelijkheid in onderwijskansen kan dus vergroot worden door een mindere mate van differentiatie of latere timing.

Mogelijkheden voor doorstroom

Latere doorstroommogelijkheden kunnen mogelijk de nadelige effecten van vroege selectie ongedaan maken. De interactie tussen vroege selectie en de openheid van een onderwijssysteem kan mogelijk bijdragen aan de onderwijsprestaties van een land. Leerlingen die in een verkeerd onderwijstype geselecteerd zijn, kunnen dan immers makkelijker van niveau veranderen. Hier staat tegenover dat doorstroommogelijkheden extra kosten met zich meebrengen. De invloed van latere doorstroommogelijkheden is niet onderzocht in de empirische literatuur.

Samenvattend wijzen deze empirische studies erop dat een onderwijssysteem met vroege selectie van leerlingen de ongelijkheid in onderwijsprestaties vergroot. Hoewel enkele studies positieve effecten op toetsresultaten laten zien, bestaat er geen eenduidig beeld van de effecten van vroege selectie op de kwaliteit van het onderwijs. Voor Nederland is gevonden dat uitstel van selectie de deelname aan en het afronden van hoger onderwijs kan vergroten. Daarmee lijkt voor Nederland de vraag aan de orde hoe mogelijke nadelige effecten van vroege selectie het best kunnen worden tegengegaan, hetzij door uitstel van selectie van leerlingen, hetzij door het vergroten van latere doorstroommogelijkheden.

3.4 Leraren en schoolleiders

Leraren spelen een sleutelrol in het onderwijs. Empirisch onderzoek laat zien dat de kwaliteit van leraren een sterk effect heeft op de prestaties van leerlingen. Verbetering van de kwaliteit van docenten kan daarom leiden tot betere prestaties van leerlingen. De empirische literatuur biedt echter weinig houvast over de determinanten van de kwaliteit van leraren. De invloed van de initiële opleiding van de docent op de prestaties van leerlingen is niet duidelijk, empirische studies geven geen consistent beeld. Ook de prestaties die leraren zelf haalden tijdens hun onderwijsloopbaan zijn geen goede voorspeller van hun bijdrage aan de prestaties van leerlingen. Training van werkende leraren en begeleiding van startende leraren door een mentor kan wel effect hebben. De ervaring van de docent blijkt slechts beperkt van invloed te zijn op de

prestaties. Alleen docenten in de eerste twee jaren van hun beroepsloopbaan dragen minder bij aan de prestaties van leerlingen dan docenten met meer ervaring. Na twee jaar in het onderwijs blijkt ervaring nog maar weinig verschil te maken. De beloningsstructuur in het onderwijs is vooral gebaseerd op ervaring en veel minder op prestaties. In verschillende landen wordt geëxperimenteerd met verschillende vormen van prestatiebeloning voor leraren. Deze experimenten wijzen erop dat het relateren van de beloning van leraren aan prestaties positief bijdraagt aan de prestaties van leerlingen. In een recente studie zijn ook positieve effecten gevonden van prestatiebeloning voor schoolleiders. Het empirisch onderzoek over leraren gaat vooral over de invloed op *de kwaliteit van het onderwijs* en levert weinig inzicht in de effecten op *de toegankelijkheid* en *de doelmatigheid*. Wel lijkt prestatiebeloning in het algemeen doelmatiger dan het verhogen van salarissen.

Voor het Nederlandse onderwijsstelsel roept dit empirische onderzoek de volgende vragen op:

1. Allereerst spelen er vragen over de aansluiting tussen opleiding en beroep. De empirische literatuur vindt geen sterke relatie tussen de initiële opleiding van de docent en de prestaties van leerlingen. Wat betekent dit voor de werving van kwalitatief goede leraren en voor de inrichting van de lerarenopleiding? Zijn opleidingseisen nodig of vormen ze juist een obstakel voor het aantrekken van goed personeel door het beperken van het aanbod van potentiële kandidaten?
2. De beloning van leraren is vooral gebaseerd op ervaring. Ervaring blijkt echter alleen in de eerste twee jaren van de beroepsloopbaan van invloed te zijn op de prestaties. Het relateren van de beloning van leraren aan de prestaties van leerlingen blijkt de kwaliteit van het onderwijs te verhogen. Wat zijn de mogelijkheden voor en de effecten van prestatiebeloning in Nederland?

Determinanten van de kwaliteit van leraren

Een verbetering van de kwaliteit van docenten kan leiden tot betere prestaties van leerlingen. Stevig bewijs hiervoor wordt geleverd in een recente studie waarin een half miljoen leerlingen in de Amerikaanse staat Texas meerdere jaren wordt gevolgd (Rivkin et al, 2005). Door de herhaalde observaties van dezelfde leerlingen is het mogelijk om heel precies de bijdrage van de docent en van de school vast te stellen. Daaruit komt naar voren dat er grote verschillen zijn in de bijdragen van docenten aan de prestaties van leerlingen. De leerkracht beïnvloedt de leerling zeer direct (zie ook de overzichtsstudies van Scheerens, 2005 en Seidel & Shavelson, 2007).

Het vaststellen van de invloed van bepaalde kenmerken van leraren op de prestaties van leerlingen is lastig vanwege selectieprocessen van leraren en scholen. Hierdoor kunnen niet geobserveerde kenmerken de schattingsresultaten van bepaalde kenmerken vertekenen. In de literatuur richt de aandacht zich op vijf kenmerken die mogelijk belangrijk kunnen zijn voor de kwaliteit van leraren: de initiële opleiding, toetsscores (prestaties in het onderwijs),

training, ervaring, salaris. Bij het kenmerk salaris wordt onderscheid gemaakt tussen een verhoging van het salaris zonder link met prestaties en met een link met prestaties.

De initiële opleiding van leraren

Veel onderzoek is gedaan naar de invloed van de initiële opleiding van leraren op hun prestaties. Dit onderzoek geeft echter geen eenduidig beeld.

Hanushek en Rivkin (2006) geven een overzicht van studies in de VS naar de effecten van de opleiding van leraren op de onderwijsprestaties van leerlingen. In 86% van in totaal 170 studies wordt geen significant effect van gevonden van de initiële opleiding van leraren. Wanneer de auteurs zich alleen richten op de studies met een hoge methodologische kwaliteit wordt in 91% van de 34 studies een insignificant effect gevonden.

In een studie van Goldhaber en Brewer (1997) worden paneldata gebruikt om te corrigeren voor niet-waargenomen kenmerken. Zij vinden een positief effect van het diploma van de docent op de prestaties van de leerlingen. Hanushek et al. (1998) gebruiken een vergelijkbare aanpak en vinden daarentegen geen effect van de initiële opleiding van de docent op de prestaties van leerlingen.

Clotfelter et al. (2008) maken gebruik van examenresultaten van verschillende vakken in North Carolina om het verband tussen behaalde diploma's van leraren en prestaties van leerlingen te onderzoeken op high-school niveau. De aanwezigheid van examenresultaten op meerdere vakken per leerling maakt het mogelijk een fixed-effects model te schatten. Deze 'within-student' schatting pakt het mogelijke endogeniteitsprobleem vanwege nonrandom toewijzing van leerlingen aan leraren aan door het effect te identificeren op variatie binnen een leerling. Ze vinden statistisch significant positieve effecten van zowel algemene diploma's als vakspecifieke diploma's op de onderwijsprestaties van leerlingen. Leerlingen met een beter gediplomeerde leraar scoren ongeveer 0.18 standaarddeviaties hoger.

Een empirische studie van Kane et al. (2008) laat echter zien dat ex ante diploma's niet de beste voorspeller zijn van de effectiviteit van leraren. Zij gebruiken longitudinale data van toetsresultaten van leerlingen op rekenen en taal over een periode van zes jaar om effecten te onderzoeken van gediplomeerde leraren in publieke scholen in New York City. Door gebruik te maken van de toetsscores in grade 4 tot grade 8 wordt de 'value-added' van leraren geschat, waarbij gecontroleerd wordt voor onder andere initiële toetsscores (in het jaar voor de tests), ervaring van de leraar en een aantal leerling-, klas-, en schoolfactoren. Ze vinden geen significante effecten van diploma status van een leraar op de onderwijsprestaties op zowel rekenen als taal. Ook Boyd et al. (2005), die vergelijkbare data gebruiken om het effect van initiële certificering op effectiviteit van leraren in New York City te evalueren, vinden dat verschillen in certificering geen of nauwelijks effect hebben op verschillen in

leerlingprestaties.⁷ Cantrell et al. (2008) onderzoeken de effecten van gecertificeerde leraren op prestaties van leerlingen. Het betreft het zogenoemde National Board for Professional Teaching Standards (NBPTS) certificaat. Dit certificaat kan worden verkregen aan de hand van een beoordeling van leraren op basis van video's en een aantal essays die door de indienende leraren moeten worden ingezonden. De auteurs vinden dat leerlingen die at random worden toegewezen aan hoogscorende aanvragers van een NBPTS-certificaat beter scoren dan studenten die een niet-indiener krijgen toegewezen, terwijl studenten die worden toegewezen aan een slecht beoordeelde aanvrager slechter scoren. Het verschil tussen slecht scorende en goed scorende aanvragers is statistisch significant. De wijze van certificering kan hiermee een voorspellende waarde hebben voor de prestaties van leerlingen.

Toetsscores

Een mogelijke indicator van de kwaliteit van leraren zijn hun eigen prestaties (cijfers) of toetsscores gedurende hun onderwijsloopbaan. Hansuhek en Rivkin (2006) geven een overzicht van studies naar de effecten van toetsscores van leraren op de prestaties van leerlingen. De belangrijkste conclusie uit dit overzicht is dat het beeld niet eenduidig is. In 22% van in totaal 41 studies wordt een significant positief effect gevonden, in 10% van de studies wordt een significant negatief effect gevonden. Wanneer alleen de meest overtuigende studies wat betreft methodologie ('value-added' schattingen waarbij de steekproef bestaat uit individuele leerlingen binnen een regio of staat) in beschouwing worden genomen, wordt in 2 van de 9 studies een positief significant effect gevonden. Eén studie vindt een significant negatief effect en 6 vinden een insignificant effect. Opgemerkt moet worden dat de toetsen sterk verschillen wat betreft inhoud en focus tussen de verschillende analyses, waardoor de gevonden effecten een mix vormen van verschillende zaken. Daarnaast blijkt dat, ook al worden er significante effecten gevonden, de toetsscores maar een klein deel in de totale variatie van effectiviteit van leraren verklaren.

Een mogelijk aanknopingspunt voor toekomstig onderzoek vormt de studie van Wayne en Young (2003), die suggereren dat toetsscores van leraren vooral gerelateerd zijn aan leerlingprestaties op specifieke vakken. Zo zouden toetsscores bijvoorbeeld meer van belang kunnen zijn voor wiskunde op het voortgezet onderwijs dan voor lezen op het primair onderwijs. Een ander waardevol onderwerp voor toekomstig onderzoek is het verschil in effect van toetsscores op verschillende groepen leerlingen. Zo suggereren een aantal studies dat toetsscores geen uniform effect hebben, maar vooral effect sorteren aan de bovenkant van de verdeling.

⁷ Zowel Kane et al. (2008) als Boyd et al. (2005) maken gebruik van nieuwe wetgeving in New York, waardoor minder strikt wordt omgegaan met vereiste certificering en meer mensen met uiteenlopende achtergrond voor de klas mogen staan. Waar Kane et al. (2008) zich vooral richten op variatie binnen groepen met een bepaalde certificering, concentreren Boyd et al. (2005) zich vooral op verschillen tussen groepen.

Training van werkende leraren en begeleiding van startende leraren

Een opleiding of training kan bijdragen aan de kwaliteit van leraren. In een economische studie van Angrist en Lavy (2001) is gevonden dat scholing van leraren een doeltreffend instrument kan zijn om onderwijsprestaties te verbeteren. Zij schatten de effecten van training van leraren op prestaties van leerlingen in het basisonderwijs in Jeruzalem. Ze maken hierbij gebruik van een interventie waarbij scholen in bepaalde buurten, met een hoger percentage immigranten en relatief lage prestatieniveaus, een budgetverhoging kregen. Deze extra financiering werd vooral gebruikt voor trainingen, waarbij leraren wekelijks door externe instructeurs les kregen om instructietechnieken voor taal en wiskunde te verbeteren. Ze vergelijken toetsresultaten op wiskunde en taal tussen 'treated' leerlingen (zij die les kregen van leraren die training hadden gevolgd) en 'control' leerlingen (zij die geen les kregen van deze leraren). Om te controleren voor (ongeobserveerde) verschillen tussen beide groepen leerlingen, maken zij gebruik van een difference-in-differences analyse, waarbij ze zowel pre-treatment toetsscores als post-treatment toetsscores van leerlingen gebruiken. Zowel deze difference-in-differences analyses, als regressie- en matching schattingen wijzen erop dat training de toetsscores op lezen en wiskunde verbetert.

Jacob en Lefgren (2004) proberen het effect van scholing te identificeren door gebruik te maken van een regeling in Chicago die slecht presterende scholen tijdelijk extra middelen toekent. Scholen in Chicago, waarin minder dan 15% van de leerlingen van de leerlingen voldeden aan de landelijke normen voor leesvaardigheid, kregen tijdelijk extra middelen die onder meer bestemd waren voor scholing van leraren. Leraren op scholen waarvan meer dan 15% aan de norm voldeed, kregen deze scholing niet. Deze regel zorgde ervoor dat 71 van 489 basisscholen in het publieke onderwijssysteem in Chicago recht hadden op extra middelen. Jacob en Lefgren (2004) gebruiken deze nonlineaire afkapgrens om het causale effect van scholing van leraren op onderwijsprestaties van leerlingen te identificeren met behulp van een regression-discontinuity design. Hiertoe werden prestaties op scholen net onder de afkapgrens vergeleken met prestaties op scholen net boven de afkapgrens. Het hoofdresultaat van de studie is dat investeringen in ontwikkeling van personeel de onderwijsprestaties van leerlingen niet hebben verbeterd. De auteurs merken op dat dit mogelijk komt door de relatief beperkte grootte van de investering.

Rockoff (2008) kijkt naar de effecten van mentoring van beginnende leraren in New York City op het vertrek uit het onderwijs en op de leerprestaties van leerlingen. Door gebruik te maken van een zogenoemde difference-in-difference methode vindt Rockoff dat mentoring een positieve invloed heeft op het afmaken van het schooljaar. Mentoring kan hiermee het behoud van startende leraren in het onderwijs positief beïnvloeden. Instrumentele variabele technieken laten zien dat extra beschikbare uren mentoring voor leraren leidt tot significant betere prestaties in rekenen en lezen.

Ervaring en andere kenmerken

De kwaliteit van de leraar kan mogelijk ook toenemen door ervaring. Hanushek et al. (1998) vinden alleen een effect van ervaring in het eerste en tweede jaar dat docenten lesgeven. Voor leraren met een langere anciënniteit wordt geen toename op onderwijsprestaties waargenomen. Kane et al. (2008) vinden vergelijkbare effecten. De effectiviteit van leraren verbetert alleen in de eerste jaren dat zij lesgeven. Zij suggereren dat onderwijsprestaties in een klas gedurende de eerste twee jaar van een leraar (ex post) een betrouwbaarder indicator is voor toekomstige effectiviteit van de leraar dan de (ex ante) certification status. Dit impliceert dat scholen beter hun energie kunnen steken in ex-post selectie, door leraren te behouden die in de eerste jaren van hun carrière goede resultaten laten zien.

Karakteristieken van leraren die latere effectiviteit van leraren voorspellen zijn van groot belang op het moment dat scholen nieuwe leraren gaan werven. Rockoff et al. (2008) concluderen dat het empirisch onderzoek weinig houvast biedt over de determinanten van de kwaliteit van leraren. Ervaring van een leraar is weliswaar in vrijwel alle empirische papers gerelateerd aan de effectiviteit, maar een wervingsbeleid dat erop gericht is alleen ervaren leraren binnen te halen is niet haalbaar. Tegelijkertijd is het belang van kwalitatief goede leraren aangetoond in empirisch werk waaruit substantiële verschillen in prestatiegroei naar voren kwamen tussen leerlingen die werden toegewezen aan verschillende leraren (zie bijvoorbeeld Rivkin et al. (2005) en Aarensen et al. (2007)). Samen hebben deze bevindingen geleid tot suggesties dat scholen meer aandacht zouden moeten hebben voor prestaties aan het begin van de loopbaan van een leraar, in plaats van veel middelen in te zetten voor werving en selectie (Gordon et al., 2006).

Rockoff et al. (2008) onderzoeken nieuwe factoren die de effectiviteit van een leraar kunnen voorspellen. Hiertoe hebben ze uitgebreid informatie verzameld bij nieuwe wiskunde leraren in het schooljaar 2006-2007 in New York over onder andere vakkennis, cognitieve vaardigheden, verschillende persoonlijkheidskenmerken en overtuigingen. Voor enkele van deze kenmerken worden statistisch significante, maar kleine effecten op onderwijsprestaties gevonden. Geen van deze individuele kenmerken kan een groot verschil in effectiviteit verklaren onder de respondenten. Wanneer ze deze kenmerken combineren met behulp van een factoranalyse tot twee factoren die de cognitieve en non-cognitieve vaardigheden van de leraren samenvatten worden economische betekenisvolle en statistisch significante effecten gevonden. De schattingen laten zien dat leerlingen die zijn toegewezen aan een leraar die één standaarddeviatie hoger scoort op ofwel de cognitieve factor ofwel de non-cognitieve factor, gemiddeld 0.03 standaarddeviatie hoger scoren op toetsen. De auteurs concluderen dat, hoewel er geen unieke factor lijkt te zijn die succes in lesgeven kan voorspellen, het gebruik van een brede verzameling aan kenmerken kan helpen bij het werven van kwalitatief goede leraren.

Salaris

Het salaris van leraren is ook een potentiële determinant van de kwaliteit van leraren. Hanushek en Rivkin (2006) presenteren een overzicht van studies naar de effecten van het salaris van leraren op onderwijsprestaties van leerlingen. In het algemeen concluderen zij dat salarissen meer positief samenhangen met leerling-prestaties dan negatief. Echter, slechts een klein deel van de studies laat significante effecten zien: in 73% van in totaal 118 studies wordt een statistisch insignificant effect gevonden. Het beeld verandert niet als alleen naar een specifieke verzameling van meer overtuigende studies wordt gekeken. Dat wil zeggen, studies waarin een toegevoegde waarde bepaald kan worden door de beschikbaarheid van voor- en nametingen en/of waarin de analyse binnen een bepaalde regio plaatsvindt. Ook voor deze groep wordt in de overgrote meerderheid (82%) van in totaal 17 studies geen significant effect gevonden. De studies die significante effecten vinden, laten een positief verband zien tussen loon en prestatie. Dit zijn echter allemaal studies binnen een bepaalde regio en betreffen dus alleen effecten van een verandering langs de looncurve en niet van een verandering van de hele curve. De auteurs merken op dat er nog andere zaken spelen die het moeilijk maken om het effect van salarissen op prestaties te identificeren. Zo is het mogelijk dat variaties in nominale salarissen voor een deel slechts verschillen reflecteren in de compensatie voor kosten van levensonderhoud, de aantrekkelijkheid van bepaalde scholen of werkomstandigheden. De meeste studies naar effecten van salarissen op onderwijsprestaties nemen dit niet mee. Een uitzondering hierop is Loeb en Page (2000), die op basis van panel data beargumenteren dat een gebrek aan goede controles voor niet-salarisaspecten en marktverschillen tussen verschillende beroepsmogelijkheden de effecten van salarissen op onderwijsuitkomstmaten in veel studies gemaskeerd hebben. Wanneer zij corrigeren voor arbeidsmarktfactoren vinden zij dat een verhoging van het salaris met 10% de uitvalpercentages in het voortgezet onderwijs met 3 tot 4% doet afnemen.

Een ander punt is het feit dat salarissen effect hebben op de beroepskeuze, de keuze van de eerste baan en mobiliteit tussen banen. Aangezien vrijwel alle studies het effect van salaris op de effectiviteit van de huidige verzameling leraren analyseert, worden effecten op de verzameling (en verdeling) van leraren niet meegenomen. Een uitzondering is Hanushek et al. (2005) die een steekproef gebruiken van leraren die van regio veranderen om het verband tussen salaris en kwaliteit van lesgeven te identificeren. Hierbij vinden zij niet dat hogere salarissen effectievere leraren aantrekt.

Meer middelen voor leraren

De kwaliteit van leraren kan mogelijk ook verbeteren door meer middelen voor leraren beschikbaar te maken voor scholen. Leuven et al. (2007) onderzoeken het effect van extra geld voor leraren op prestaties van leerlingen in Nederland. Scholen met ten minste 70% achterstandsleerlingen ontvingen additionele financiering voor personeel. De discontinuïteit die ontstaat door deze regel wordt benut om het effect van subsidies voor personeel te identificeren.

Hiertoe worden door prestaties van leerlingen op scholen net boven de afkapgrens vergeleken met die van leerlingen op scholen net onder de afkapgrens. Ze vinden geen effecten van de personeelssubsidies op de toetsscores van leerlingen.

Samenvattend kan worden geconcludeerd dat het lastig is om effecten van salaris van leraren op onderwijsprestaties te identificeren. Uit de bestaande studies volgt geen consistent beeld van de invloed van de hoogte van het salaris op prestaties.

Prestatiebeloning

Prestatiebeloning voor leraren kan leiden tot betere onderwijsprestaties van leerlingen via betere prikkels. De beloning kan gekoppeld zijn aan de individuele prestaties (individuele prestatiebeloning) of aan de prestaties van een groep leraren, zoals een afdeling of school (teambeloning). De prestatiebeloning zorgt er in het laatste geval voor dat managers en werknemers dezelfde doelen nastreven. Prestatiebeloning in het onderwijs is een controversieel onderwerp. Problemen rond het meten van de 'output' van scholen of individuele docenten maken het gebruik van prestatiebeloning lastiger dan in sommige andere sectoren. Zo is het moeilijk om de bijdrage van de leraar of school aan de onderwijsprestaties precies vast te stellen, aangezien de toegevoegde waarde (het verschil in toetsresultaten tussen twee meetmomenten) ook wordt bepaald door inputs van het gezin en de klasgenoten. Door de toegevoegde waarde alleen aan de school toe te schrijven, kunnen prestaties van scholen verkeerd ingeschat worden (Todd en Wolpin, 2003). Ook kunnen onderwijsprestaties worden gemanipuleerd door bijvoorbeeld zwakke leerlingen buiten de tests te houden (Ladd en Walsh, 2002; Jacob, 2002; Figlio en Getzler, 2002) of door *teaching-to-the-test* (Jacob, 2002). Naast problemen met het meten van onderwijsprestaties zitten er ook in algemene zin risico's verbonden aan het gebruik van prestatiebeloning. Zo kan individuele prestatiebeloning ertoe leiden dat collega's minder goed gaan samenwerken, of elkaar zelfs gaan tegenwerken (Lazear en Rosen, 1981). Teambeloning kan dit probleem van non-coöperatief gedrag verhelpen, maar daar staat tegenover dat een beloning gebaseerd aan een groepsprestatie kan leiden tot liftersgedrag (Lazear, 1999). Er bestaat bovendien enige evidentie dat prestatiebeloning nadelig kan uitpakken wanneer werknemers intrinsiek gemotiveerd zijn (Kreps, 2007; Frey en Jegen, 2001).

De laatste jaren zijn veel nieuwe beleidsprogramma's met prestatieprikkels in het onderwijs gestart, met name in de VS en Engeland. De empirische evaluaties wijzen op positieve effecten van deze programma's op de prestaties van leerlingen. Hieronder worden de beschikbare evaluaties voor zowel teambeloningen als individuele prestatiebeloningen samengevat.

Teambeloning

Lavy (2002a) evalueert de effecten van een teambeloning in het Israëlische voortgezet onderwijs met een regression-discontinuity design. In 1996 kreeg een deel van de Israëlische scholen een financiële beloning gebaseerd op onderwijsprestaties en uitvalpercentage in het laatste jaar van het voortgezet onderwijs. Een groot deel van de beloning werd verdeeld onder de leraren (proportioneel naar bruto salaris) en de rest werd gebruikt voor de verbetering van faciliteiten op school. Het selectiecriteria voor deelname was dat een school de enige in zijn soort moest zijn in de regio. Als controlegroep zijn alle scholen gekozen waarvan er twee soorten in hun regio waren. Dit zijn de scholen die net boven de afkapgrens voor deelname vallen. De resultaten suggereren dat prestatiebeloning voor leraren tot een significante verbetering van onderwijsprestaties leidt, zoals hogere gemiddelde toetsscores, een toename in het behaalde aantal studiepunten en een afname in het aantal voortijdig schoolverlaters. De auteurs vinden daarnaast dat ook een alternatief beloningsprogramma, waarbij scholen meer middelen kregen voor extra lestijd voor zwakkere leerlingen, positieve effecten heeft op onderwijsprestaties van leerlingen. Een vergelijking tussen de twee programma's suggereert echter dat het eerste programma, gericht op directe beloning van leraren, kosteneffectiever is.

Ladd (1999) evalueert een teambeloningsprogramma dat in 1991 in Dallas is ingevoerd. Hiertoe wordt gebruikt gemaakt van een panel data set voor scholen in grote Texaanse steden om de relatieve verbetering in onderwijsprestaties van leerlingen in Dallas ten opzichte van de andere steden te bestuderen. De resultaten laten zien dat de slagingspercentages in Dallas meer zijn toegenomen dan in andere grote Texaanse steden, wanneer gecontroleerd wordt voor stad- en schoolspecifieke verschillen. Verder blijkt dat de impact sterk kan verschillen tussen etnische groepen, waarbij vooral voor Latijns-Amerikaanse en blanke leerlingen de meest positieve resultaten gevonden worden. Daarnaast vindt Ladd ook dat de teambeloning leidt tot minder voortijdig schoolverlaters en meer verloop onder schooldirecteuren.⁸

Glewwe et al. (2003) evalueren een gerandomiseerd experiment in Kenia waarbij leraren op basisscholen een beloning konden krijgen gebaseerd op toetsresultaten van leerlingen. Ze vinden dat leerlingen op de scholen met beloning hogere toetsscores haalden. Voor enkele van de toetsen waren de verschillen in scores statistisch significant. Een verder onderzoek naar het onderliggende mechanisme suggereert echter dat dit niet voortkwam uit meer aanwezigheid van leraren, verandering in huiswerkopgaven of pedagogiek, maar door meer voorbereidingsessies voor de specifieke toetsen. Gedurende het programma scoorden leerlingen op de interventiescholen weliswaar hoger, maar dit effect bleef niet langer te bestaan na afloop van

⁸ Een mogelijk probleem in de interpretatie van resultaten van Ladd (1999) is dat er voor de invoering van het beloningsprogramma al sprake was van positieve effecten, veroorzaakt door publiciteit rondom de hervormingsplannen. Een dergelijk 'Hawthorne' effect zou na verloop van tijd moeten verdwijnen, maar Ladd vindt een blijvend positief effect op studieprestaties.

het programma. Dit is volgens de auteurs in lijn met de hypothese dat leraren zich vooral richtten op het beïnvloeden van de toetscores op korte termijn.

Individuele beloning

Lavy (2003) evalueert de effecten van een bonusprogramma voor leraren in Israël. Alleen scholen waarvan minder dan 45% van de leerlingen slaagde voor het examen van het voortgezet onderwijs (matriculation exam) werden geselecteerd voor deelname. Om het causale effect van de beloning op studieprestaties vast te stellen gebruikt Lavy drie identificatiemethoden.

Allereerst wordt gebruik gemaakt van een klassiek (sharp) regression-discontinuity design gebaseerd op de selectieregel voor scholen. Hierbij worden scholen met een matriculation rate net onder de 45% vergeleken met een controlegroep van scholen met een matriculation rate net boven de 45%. De tweede methode betreft een natuurlijk experiment waarbij gebruik wordt gemaakt van willekeurige meetfouten in de toewijzingsvariabele, die ontstaan omdat scholen worden toegewezen op basis van voorlopige informatie. Ten slotte gebruikt Lavy propensity score matching, waarbij een controlegroep wordt geconstrueerd die sterk vergelijkbaar is op basis van observeerbare kenmerken. De resultaten van de verschillende methoden zijn goed vergelijkbaar. Het programma leidt tot een significante verbetering van onderwijsprestaties. Zo verbeteren de resultaten op wiskunde en Engels met respectievelijk 19 en 21%.

Atkinson et al. (2004) vinden ook positieve effecten van een prestatiebeloningsprogramma in Engeland. In 1999 heeft de Engelse regering een beloningsprogramma voor leraren geïntroduceerd, waarbij de vooruitgang in leerlingprestaties één van de belangrijkste criteria was. Atkinson et al. (2004) verzamelden longitudinale data op leraarniveau zowel voor als na de introductie van dit programma. Ze gebruiken een difference-in-differences analyse waarbij wordt gecontroleerd voor individuele effecten op leerling-, leraar- en schoolniveau. De resultaten laten zien dat het programma de gemiddelde toetscores op examens verbeterden met een ongeveer een half punt, wat overeenkomt met 0.73 standaarddeviatie.

Eberts et al. (2002) analyseren het effect van een individuele prestatiebeloning voor leraren op een middelbare school in Michigan. Leraren konden een financiële bonus verdienen wanneer zij een lage schooluitval realiseerden of wanneer zij hoog scoorden op een evaluatie door studenten. In de analyse worden verschillende prestaties van leerlingen op de betrokken school vergeleken met prestaties van leerlingen op een vergelijkbare middelbare school waar een meer traditioneel beloningssysteem werd gebruikt. Een difference-in-differences analyse laat zien dat de prestatiebeloning leidt tot lagere uitvalpercentages. Er worden echter geen positieve effecten gevonden op gemiddelde studieprestaties, gemiddelde aanwezigheid en slagingspercentage. Op deze uitkomstvariabelen scoorde de school met prestatiebeloning zelfs slechter. Deze resultaten kunnen volgens de auteurs verklaard worden doordat de prikkel zich voornamelijk richt op het

verminderen van uitval van zwakkere leerlingen.⁹ Eberts et al. (2002) concluderen dat prestatieprikkels leraren kunnen motiveren om uitkomsten te bereiken die direct beloond worden, maar wijzen ook op het gevaar dat dit kan leiden tot onbedoelde en ongewenste resultaten wanneer een beloningsprogramma niet zorgvuldig geconstrueerd en ingevoerd wordt.

Muralidharan en Sundararamen (2008) evalueren een gerandomiseerd experiment waarbij een beloningsprogramma voor leraren werd ingevoerd in India. Het programma bestond uit financiële bonussen voor leraren gebaseerd op de gemiddelde verbetering van prestaties van leerlingen op toetsscores. Twee jaar na de invoering van het programma scoorden leerlingen op de interventiescholen significant beter op toetsscores dan leerlingen op de controlescholen. Het effect bedroeg respectievelijk 0.28 en 0.16 standaarddeviaties op wiskunde- en taaltoetsen. Daarnaast vergelijken de auteurs scholen met een individueel beloningsprogramma met scholen met een groepsbeloningsprogramma. Het blijkt dat beide programma's in het eerste jaar even goed werken, maar dat de individuele beloningen voor significant betere leerlingprestaties zorgen in het tweede jaar. Ook presteren de scholen met een individueel beloningsprogramma beter dan andere aselekt gekozen basisscholen die additionele school inputs ontvingen ter waarde van de bonussen.

Schoolleiders

Naast positieve effecten van prestatieprikkels voor leraren, blijkt prestatiebeloning voor schoolleiders ook effectief. Lavy (2008) gebruikt de gefaseerde invoer van een salarisverhoging van 50% voor schoolleiders om te onderzoeken wat de effecten op leerresultaten zijn. Op scholen waar de beloning eerder is ingevoegd zijn de resultaten hoger dan op de scholen waar het 1 of 2 jaar later is ingevoerd. De toename is weliswaar gering (5 tot 10% stijging), maar gezien de relatief lage kosten concludeert Lavy dat prestatieprikkels voor schoolleiders bijdraagt aan effectiever onderwijs.

Samenvattend kan worden geconcludeerd dat de empirische studies wijzen op positieve effecten van prestatieprikkels voor leraren. Prestatiebeloning kan echter ook ongewenste neveneffecten hebben, zoals *teaching-to-the-test*. Het toepassen van prestatieprikkels in het onderwijs is daarom gecompliceerd en vereist maatwerk. Gezien de potentiële positieve effecten van prikkels, en het belang van zorgvuldig geconstrueerde en ingevoerde beloningsprogramma's om ongewenste effecten te voorkomen, is de werking van prestatieprikkels in het onderwijs een belangrijk onderwerp voor toekomstig onderzoek.

⁹ Een punt van zorg in deze analyse is wel dat er belangrijke (significante) verschillen bestaan in gemiddelde cijfers en daarmee in vergelijkbaarheid van de scholen. Deze verschillen zullen niet wegvallen uit de difference-in-differences analyse als er een non-lineair verband is tussen deze kenmerken en de effectiviteit. Dit laat ruimte voor alternatieve verklaringen voor het gevonden verschil in ontwikkeling van studieprestaties op deze scholen.

3.5 Speciaal onderwijs

Vrijwel alle landen kennen naast het reguliere onderwijs vormen van speciaal onderwijs voor leerlingen met een bepaalde handicap. Daarbij kunnen verschillen optreden ten aanzien van de verhouding regulier/speciaal (wat is het aandeel leerlingen speciaal onderwijs binnen de totale leerlingenpopulatie) en de toewijzing aan het speciaal onderwijs. Ook worden leerlingen uit het speciaal onderwijs soms opgevangen of herplaatst in het reguliere onderwijs. De empirische kennis over deze stelselkenmerken en hun effecten op *kwaliteit, doelmatigheid en toegankelijkheid* is erg beperkt. Tussen landen bestaat aanzienlijke variatie in aandelen leerlingen in het speciaal onderwijs. Het effect daarvan op de totale onderwijsuitkomsten is niet bekend. De empirische literatuur geeft aanwijzingen dat de vraag naar speciaal onderwijs gevoelig is voor de financieringsystematiek. Voor de VS is het oorzakelijk effect geschat van een programma van het speciaal onderwijs. Het programma bleek positief bij te dragen aan de prestaties in rekenen en in taal. De effecten van het opvangen van leerlingen uit het speciaal onderwijs in het regulier onderwijs (voor henzelf en hun klasgenoten) zijn niet bekend. Mede gezien de substantieel hogere kosten van het speciaal onderwijs vergeleken met het regulier onderwijs lijkt verder onderzoek naar de effectiviteit van verschillende vormen van speciaal onderwijs in Nederland, alsmede van het opvangen van deze leerlingen in het regulier onderwijs, zeer gewenst.

Leerlingen in het speciaal onderwijs

In het Nederlandse basisonderwijs wordt onderscheid gemaakt tussen drie soorten zorgleerlingen: leerlingen op het speciaal onderwijs, zogenaamde 'rugsakleerlingen' op het reguliere onderwijs en leerlingen op het speciaal basisonderwijs. Leerlingen op het speciaal onderwijs en rugsakleerlingen hebben ernstige lichamelijke of geestelijke handicaps of ernstige gedragsproblemen. Rugsakleerlingen zijn leerlingen met een indicatie voor het speciaal onderwijs die in het regulier onderwijs worden opgevangen. Leerlingen op het speciaal basisonderwijs hebben lichtere handicaps, zoals ADHD, dyslexie, lichte vormen van autisme of ze zijn hoogbegaafd.¹⁰

In Nederland is de laatste jaren sprake van een toename van leerlingen in de relatief dure vormen van speciaal onderwijs, te weten in het speciaal onderwijs en rugsakleerlingen. Gezien deze ontwikkeling is een aantal stelselkenmerken van belang. Het eerste is het *aandeel zorgleerlingen in het totaal aantal leerlingen*. Een internationale vergelijking geeft informatie over verschillen van opvatting tussen landen over de criteria wanneer een kind een extra zorg

¹⁰ De OECD (2004) onderscheidt ook een 'categorie C' leerling, die extra zorg krijgen om een sociaaleconomische reden. In Nederland valt hier in elk geval de gewichtenregeling van het onderwijsachterstandenbeleid onder. We beschouwen het onderwijsachterstandenbeleid niet als stelselkenmerk dat actueel is voor beleid. Daarom laten we deze categorie hier buiten beschouwing. We verwaarlozen kinderen die zijn vrijgesteld van leerplicht om dezelfde reden.

bij onderwijs nodig heeft.¹¹ Leerlingen op het speciaal onderwijs vallen bij de OECD (2004) in de 'categorie A'; leerlingen op het speciaal basisonderwijs en het praktijkonderwijs (van het voortgezet onderwijs) vallen in 'categorie B' (OECD, 2004, blz. 153). Het blijkt dat Nederland op het gemiddelde zit met het aandeel categorie A leerlingen en bij categorie B leerlingen er niet uitspringt. Wel zitten in Nederland veel zorgleerlingen op aparte scholen en minder op reguliere scholen vergeleken met andere landen (blz. 84 en 89). Echter, deze cijfers hebben betrekking op het pre-rugzaktijdperk.

Het tweede stelselkenmerk is het *functioneren van de indicatieorganen* die de toelating tot het speciaal onderwijs en het speciaal basisonderwijs regelen. Hierover zijn voor zover bekend geen cijfers beschikbaar.

Het derde stelselkenmerk is de opvang van kinderen met handicaps op het reguliere onderwijs. Vanaf 2003 hebben leerlingen met een indicatie voor het speciaal onderwijs in Nederland ook de mogelijkheid op hun reguliere school te blijven als *rugzakleerling*. De reguliere school ontvangt voor zijn extra onderwijs het geld dat in de rugzak zit. Het aantal rugzakleerlingen is sinds 2003 sterk gestegen. De OECD (2004) geeft geen gegevens over rugzakleerlingen van andere landen. In Nederland was hun aantal nul in 1998/99 aangezien de mogelijkheid om rugzakleerling te zijn toen niet bestond.

Onderzoek naar de effectiviteit van beleid voor het onderwijs aan zorgleerlingen wordt bemoeilijkt door verschillen tussen stelsels van het speciaal onderwijs en het regulier onderwijs enerzijds en moeilijk waarneembare verschillen in capaciteiten tussen zorgleerlingen en gewone leerlingen anderzijds. Daarnaast is er een schaarse beschikbaarheid van cijfers over de deelname aan en kosten van het onderwijs aan zorgleerlingen.

Van de *deelname* aan het speciaal onderwijs zijn alleen cijfers beschikbaar van de OECD (2004) over het schooljaar 1998-1999.¹² Hieruit blijkt dat vrijwel alle landen naast het reguliere onderwijs ook aparte vormen van onderwijs voor zorgleerlingen kennen, maar dat de stelsels internationaal zo sterk verschillen dat de gegevens moeilijk internationaal vergelijkbaar zijn. Er zijn dan ook geen internationaal vergelijkende studies naar de effectiviteit van speciaal onderwijs beschikbaar.

Een van de weinige empirische studies is Hanushek et al. (1998), waarin de effectiviteit van speciaal onderwijs in Texas onderzocht wordt. Zij maken gebruik van een database met longitudinale gegevens van 200 duizend leerlingen en concentreren zich op leerlingen die wisselen tussen speciaal onderwijs en regulier onderwijs, of wisselen in de mate van zorg op het reguliere onderwijs. Hierdoor kunnen de toenames in prestaties van individuele leerlingen op het regulier onderwijs vergeleken worden met die op het speciaal onderwijs. Ondanks het kleine

¹¹ We gaan er van uit dat de OECD-landen onderling weinig van elkaar verschillen in de prevalentie van handicaps en leerproblemen. Verschillen in deelname zijn dan te verklaren door verschillen in toetredingscriteria tot extra onderwijszorg.

¹² De OECD heeft ook cijfers verzameld voor het schooljaar 1995/1996, maar deze zijn van lagere kwaliteit en door definitieveranderingen niet goed vergelijkbaar met die van 1998/99.

percentage leerlingen dat wisselt, levert de grote database voldoende waarnemingen op voor een betrouwbare schatting. De auteurs vinden dat zorgleerlingen op het speciaal onderwijs meer vooruitgang boeken in rekenen en taal dan op het reguliere onderwijs. Het effect op rekenen is groter dan op taal. Een groter aandeel zorgleerlingen op het regulier onderwijs leidt niet tot slechtere prestaties van de gewone leerlingen. Het onderzoek gaat niet in op de vraag of de baten van de hogere toename in toetscores opwegen tegen de hogere kosten van het speciaal onderwijs.

Over de effecten van adaptief onderwijs bestaat een overzichtsstudie van Blok en Breetvelt (2002). Zij concluderen dat de invulling hiervan sterk verschilt tussen scholen, en eventuele effecten ook. Op basis van 16 studies komen ze tot de conclusie dat er slechts enkele aanwijzingen zijn dat adaptief onderwijs effect heeft. In een literatuurbespreking zijn Smeets en Rispens (2008) positiever omdat enkele aanvullende Nederlandse studies wel meer effect laten zien. Vermaas en van der Pluijm (2005) komen, op basis van een overzicht van onderzoeksrapporten en reviews tot een vergelijkbare conclusie, al zien zij ook knelpunten. Een Europese vergelijking over het speciaal onderwijs is te vinden in Meijer (2003).

Cullen en Rivkin (2001) maken gebruik van microgegevens in Texas en Chicago om te onderzoeken in hoeverre zorgleerlingen bijzondere problemen ondervinden bij een toename in schoolkeuze. Aan de ene kant ondervinden deze leerlingen mogelijk extra toetredingsbelemmeringen voor het regulier onderwijs, omdat scholen prikkels hebben om zorgleerlingen af te wijzen. Ten eerste vergt het aanbieden van onderwijs aan deze groep leerlingen extra inspanningen van de school. Ten tweede schrikt een school wellicht gewone kinderen af wanneer zij veel zorgleerlingen toelaat. Aan de andere kant kunnen zorgleerlingen mogelijk profiteren van beter onderwijs, door de grotere concurrentie tussen scholen als gevolg van meer schoolkeuze. De auteurs laten zien dat de invloed van een toename van schoolkeuze voor zorgleerlingen afhangt van de verantwoordelijkheid van de betrokken instellingen en de wijze van financiering. In hoeverre een toename van schoolkeuze voordelig is voor zorgleerlingen hangt af van de details van de inrichting van het onderwijsstelsel.

Invloed financieringssystematiek op het aandeel zorgleerlingen

Cullin (1999) vindt een relatie tussen de toename van het aantal zorgleerlingen in Texas en de financiering door de staat die lokale schooldistricten beloont om leerlingen als zorgleerling te diagnosticeren: 10% extra inkomen van een schooldistrict voor een zorgleerling door staatsfinanciering genereert 1,4% meer zorgleerlingen. Het grootste effect is bij zorgleerlingen met niet-lichamelijke handicaps, zoals spraakproblemen en emotionele problemen. Met andere woorden: hoe minder scherp de diagnose te stellen is, hoe sterker de diagnose van zorgleerling wordt bepaald door geldprikkel. Het onderzoek berust op data van ongeveer duizend schooldistricten in Texas gedurende 1991- 1996. Het causale effect is bepaald door gebruik te maken van de verandering in het financieringssysteem in 1993. Ook Sack (1998) en Monk (1990) vinden een relatie tussen financiële prikkels en de vraag naar speciaal onderwijs.

Samenvattend kan geconcludeerd worden dat de empirische literatuur over de effectiviteit van speciaal onderwijs nog zeer beperkt is. De empirische literatuur geeft aanwijzingen dat de vraag naar speciaal onderwijs gevoelig is voor de financieringsystematiek. Mede vanwege de substantieel hogere kosten van het speciaal onderwijs vergeleken met het regulier onderwijs en de recente sterke stijging in het aantal rugzakleerlingen in Nederland, lijkt verder onderzoek naar de effectiviteit van verschillende vormen van onderwijs voor zorgleerlingen zeer waardevol.

3.6 Achterstandenbeleid

Nederland kent verschillende vormen van achterstandenbeleid. Zo zijn in de afgelopen jaren meer middelen ingezet voor programma's voor voor- en vroegschoolse educatie (VVE) met als doel het verminderen van onderwijsachterstanden bij risicogroepen. Daarnaast ontvangen scholen in het Nederlandse basisonderwijs extra middelen voor leerlingen met een lagere sociaaleconomische achtergrond. Ook worden de laatste jaren meer middelen ingezet om het voortijdig schoolverlaten tegen te gaan.

Voor kwalitatief goede VVE-programma's zijn in de Verenigde Staten positieve effecten gevonden bij risicoleerlingen. Evaluaties op basis van gecontroleerd experimenten hebben laten zien dat deze daadwerkelijk geleid hebben tot een verbetering van het onderwijsniveau en de latere arbeidsmarktpositie van de deelnemers. Daarbij bleek dat er ook substantiële baten waren voor de rest van de samenleving in termen van vermindering van criminaliteit en inactiviteit. De vormgeving van de projecten bleek belangrijk te zijn. Intensieve projecten van minimaal vier dagdelen per week met goed opgeleide begeleiders lijken betere resultaten te geven dan minder intensieve programma's.

In het Nederlandse onderwijs, vooral in het basisonderwijs, ontvangen scholen meer middelen voor leerlingen met een lagere sociaaleconomische achtergrond. Deze zogenoemde compenserende bekostiging is niet geëvalueerd in onderzoek met een experimentele opzet. Empirisch onderzoek naar de effecten van het verhogen van de onderwijsmiddelen van scholen, zonder daaraan prestatie-eisen te verbinden, geeft een wisselend beeld. In een recente Nederlandse evaluatie, met een experimentele opzet, werd gevonden dat achterstandsscholen niet beter gingen presteren nadat zij 10 % meer middelen hadden ontvangen. In Nederland worden achterstandsmiddelen vaak gebruikt voor het verkleinen van klassen. De empirische literatuur laat zien dat kleinere klassen wel leiden tot betere prestaties maar dat dit zeer veel middelen vergt. Anders gezegd, klassenverkleining leidt wel tot meer *kwaliteit* van onderwijs maar is niet *doelmatig*.

Voortijdig schoolverlaten is een hardnekkig probleem. De empirische literatuur laat zien dat veel programma's niet werken. Daarnaast blijkt de effectiviteit van interventies vaak niet bekend te zijn. Studies met een experimentele opzet (inclusief controlegroepen) geven twee veelbelovende richtingen. In de eerste plaats blijken projecten die gebruik maken van financiële

prikkels voor leerlingen, leraren en scholen effectief te zijn. Positieve ervaringen zijn hiermee opgedaan in het Verenigd Koninkrijk, de VS, en Israël. Ook zijn veelbelovende ervaringen opgedaan met langdurige en intensieve programma's met coaches, gericht op de sociale ontwikkeling van risicjongeren.

Een algemene les uit de empirische literatuur over vermindering van onderwijsachterstanden lijkt te zijn dat vroeg ingrijpen het best werkt: voorkomen werkt beter dan genezen. Effectieve interventies voor risicogroepen dragen zowel bij aan *de kwaliteit van het onderwijs als aan de toegankelijkheid van het onderwijs*. De inzet van achterstandsmiddelen in Nederland, die vooral gericht is op VVE en het primair onderwijs, lijkt in lijn met de empirische evidentie. De vragen die nu aan de orde zijn gaan over:

1. Het vinden van de meest effectieve programma's en interventies in de Nederlandse context, aanpakken die kosteneffectiever zijn dan klassenverkleining.
 2. Het prikkelen van scholen tot het gebruik van de meest kosteneffectieve aanpakken.
- Andere belangrijke vragen, die ook eerder naar voren kwamen, gaan over schoolkeuze en segregatie, en daaraan verbonden vragen over de kwaliteit van (leraren op) scholen met veel achterstandsleerlingen.

3.6.1 Voor- en vroegschoolse educatie

Beleidsintensivering gericht op voor- en vroegschoolse opvang heeft tot doel de vroege ontwikkeling van kinderen te bevorderen en hun mogelijkheden voor vervolgonderwijs te vergroten, en daarmee uiteindelijk hun latere maatschappelijke positie en kansen op de arbeidsmarkt te versterken. Een lage of niet afgemaakte opleiding vergroot de kans op inactiviteit, maatschappelijke uitsluiting en zelfs op crimineel gedrag en kan invloed hebben op de latere gezondheid. Hieraan zijn kosten verbonden voor de samenleving, bijvoorbeeld in de vorm van sociale uitkeringen, ziektekosten, en kosten van politie, justitie en sociale hulpverlening. De potentiële baten van dit beleid zijn hoog. Onderzoek van buitenlandse projecten voor risicoleerlingen leert dat deze programma's geleid hebben tot een verbetering van het onderwijsniveau en de latere arbeidsmarktpositie van de deelnemers. Daarbij bleek dat er ook substantiële baten waren voor de rest van de samenleving in termen van vermindering van criminaliteit en inactiviteit (Currie, 2001, Heckman, 2000). De resultaten van deze projecten met kinderen uit achterstandsgroepen in de Verenigde Staten konden overtuigend worden vastgesteld omdat gebruik werd gemaakt van een experimentele opzet. Op basis van loting werd bepaald welke kinderen mochten deelnemen aan educatieve programma's en welke kinderen in een controlegroep kwamen die geen educatief programma kreeg aangeboden. De ervaringen van deze kinderen in en na het onderwijs zijn gevolgd, soms wel 30 jaar lang. Specifiek zijn er vier studies met een experimenteel design waarbij weinig uitval optrad: het Early Training project, het Carolina Abecedarianproject, het Perry Preschool project en het Milwaukee project.

Het Early Training project werd uitgevoerd in het begin van de jaren zestig in Murfreesboro, Tennessee. Het programma was gericht op 4- en 5-jarigen en bestond uit wekelijkse bezoeken aan het gezin in aanvulling op tien weken deelname aan een deeltijd voorschools traject gedurende twee of drie zomers. Vergeleken met de drie andere programma's is het Early Training project het minst intensief. Als uitkomst werd gevonden een sterke daling van de verwijzing naar het speciaal onderwijs tot 12 jaar (5 % van de experimentele groep versus 29 % in de controle groep) en, hoewel niet significant, werden hogere testcores, minder uitval en het vaker behalen van een high school diploma gevonden. Het Abcedarian Project werd uitgevoerd in de jaren zeventig en was gericht op kinderen tussen 4 maanden en 8 jaar met een relatief ongunstige sociaaleconomische achtergrond. Ook hier was sprake van 'random' toewijzing van kinderen aan het project (en aan de controlegroep). De kinderen werden fulltime opgevangen door het kinderdagverblijf gedurende het hele jaar. Bij de activiteiten met de kinderen lag de nadruk op taalontwikkeling in een rijke leeromgeving. De resultaten werden op de korte en langere termijn, op een leeftijd van 15 en 21 jaar, onderzocht. Op 15 jarige leeftijd bleken de deelnemers t.o.v. de controlegroep beter te presteren in taal en rekenen en minder vaak te doubleren of te worden verwezen naar het speciaal onderwijs. Op 21-jarige leeftijd bleek dat de deelnemers langer onderwijs bleven volgen, vaker een universitaire studie volgden, vaker een baan kregen en later begonnen aan gezinsvorming. Het project liet ook zien dat de effecten sterker waren voor kinderen die op jongere leeftijd deelname aan het project (0-5 jaar) dan voor degenen die later deelnamen. Waarschijnlijk het meeste bekende en meest uitgebreid besproken programma is het Perry-Preschool programma van de High/Scope Foundation dat in het begin van de jaren zestig is uitgevoerd in Ypsilanti in de Verenigde Staten. Kinderen met een relatief ongunstige sociaaleconomische achtergrond en met een IQ lager dan gemiddeld werden 'random' toegewezen aan een intensief één tot tweejarig programma waarbij de ouders wekelijks werden bezocht. De verhouding leerling-docent was 6 op 1 en aan de opleiding van de docent werden hoge eisen gesteld. Het unieke aan het programma is dat de deelnemers langdurig werden gevolgd; er is informatie beschikbaar over de situatie van de deelnemers bij een leeftijd van 27 jaar. Deze informatie is gebruikt om een kostenbaten analyse uit te voeren. Deelnemers aan het programma hadden hogere testcores, volgden meer onderwijs, verdienden hogere salarissen en waren minder vaak betrokken bij criminele activiteiten dan degenen in de controlegroep. Gemeten tot en met 27 jaar levert elke dollar besteed aan het programma 5.7 dollar op, projecties naar de toekomst geven een totale verwachte opbrengst van 8.7 dollar. De belangrijkste opbrengst van het programma is de reductie van crimineel gedrag (65 % van de totale opbrengst).

Het Milwaukee-project, uitgevoerd tussen 1968 en 1978, bestond uit voltijds kinderopvang en training van moeders. Het project was gericht op kinderen van 3-6 maanden en liep tot en met 5-jarigen. De resultaten wijken af van de andere projecten in de zin dat kinderen in de experimentele groep hogere IQ-scores hadden rond 14 jaar maar niet beter scoorden in het onderwijs.

Ludwig & Miller (2007) onderzoeken de effecten van Head Start, een op brede schaal ingevoerd voor- en vroegschoolse educatieprogramma in de VS.¹³ Meer dan 900.000 kinderen in de VS nemen op dit moment deel aan Head Start, waarmee ongeveer 7 miljard dollar gemoeid is. De effecten worden geschat door gebruik te maken van een grote en langdurige discontinuïteit in de financiering van Head Start over de tijd. De schattingen geven sterke en blijvende effecten van Head Start op het uiteindelijk behaalde opleidingsniveau.¹⁴ Er wordt geen bewijs gevonden voor statistisch significante blijvende effecten op toetscores in lezen en rekenen (in het achtste jaar). Het patroon van uitkomsten - wegebbende effecten op toetscores, maar lange-termijneffecten op behaald onderwijsniveau - komt overeen met eerdere evaluaties van Head Start en van Perry Preschool (zie Currie & Thomas, 1995; Garces, Thomas, and Currie 2002, Schweinhart et al. 2005). Recent onderzoek van Heckman et al. (2006) benadrukt dat zelfs beperkte toenames in cognitieve of non-cognitieve vaardigheden genoeg kunnen zijn om blijvende effecten op behaald onderwijsniveau te genereren. Een andere belangrijke bevinding is een grote afname in sterftecijfers onder kinderen aan oorzaken die *beïnvloed* zouden kunnen zijn door Head Start.

Samengevat laten de studies zien dat voorschoolse interventies positieve lange termijn effecten kunnen hebben voor kinderen in termen van betere prestaties in het onderwijs, op de arbeidsmarkt en een afname van de deelname aan criminele activiteiten. Daarbij blijkt de vormgeving van de projecten belangrijk te zijn. Intensieve projecten van minimaal vier dagdelen per week met goed opgeleide begeleiders lijken betere resultaten te geven dan minder intensieve programma's. Uiteraard kunnen deze resultaten niet één op één vertaald worden naar Nederland maar de Amerikaanse resultaten zijn zeker veelbelovend.

Wegen de baten op tegen de kosten? Een belangrijk deel van de baten van deze programma's wordt pas op langere termijn gerealiseerd, aangezien de interventie is gericht op jonge kinderen en de baten merendeels tot stand komen gedurende hun arbeidsmarktloopbaan. De buitenlandse projecten laten zien dat de maatschappelijke baten van deze projecten nog steeds opwegen tegen de kosten als rekening gehouden wordt met de lange termijn voor het beschikbaar komen van de baten. Een neveneffect van dit beleid is dat het toekomstige inkomensverschillen vermindert. Vanwege het verschil in context, bijvoorbeeld qua onderwijssysteem, is niet duidelijk of deze hoge maatschappelijke rendementen ook daadwerkelijk in Nederland zullen worden gehaald. Zo kunnen leerlingen in Nederland al vanaf 4 jaar naar de basisschool.

¹³ Behalve voor- en vroegschoolse educatie (dat 40 procent van de kosten vormt) omvat dit programma ook een aantal andere elementen op de gebieden betrokkenheid van ouders, voeding, sociale dienstverlening, mentale dienstverlening en diensten op het vlak van gezondheid.

¹⁴ Gemeten effecten zijn op het halen van een high school diploma en op deelname aan college.

3.6.2 Meer publieke middelen / compenserende bekostiging

Het Nederlandse onderwijs kent een compenserende bekostiging, vooral in het basisonderwijs. Scholen krijgen meer middelen voor leerlingen met een lagere sociaaleconomische achtergrond. Deze middelen worden toegekend onafhankelijk van de prestaties van scholen. De compenserende bekostiging in Nederland is niet geëvalueerd in onderzoek met een experimentele opzet. Empirisch onderzoek naar de effecten van het verhogen van de onderwijsmiddelen van scholen, zonder daaraan prestatie-eisen te verbinden, geeft een wisselend beeld. Woessman (2007) concludeert uit een internationale vergelijking van onderwijsuitgaven dat hogere uitgaven niet samenhangen met betere leerlingprestaties. Deze bevindingen zijn in lijn met een overzichtstudie van Hanushek (2006) naar de invloed van extra middelen voor onderwijs op de prestaties van leerlingen. Deze overzichtstudie geeft een sterk wisselend beeld, veel studies vinden geen significante effecten, en slechts een klein deel van de studies vindt significant positieve effecten. Daar staan ook enkele significante negatieve effecten tegenover. Hanushek concludeert dat een uitbreiding van de middelen voor onderwijs geen garanties geeft voor betere onderwijsprestaties. Voor Nederland wordt dit bevestigd door de hiervoor al besproken studie van Leuven et al. (2007). Deze studie laat zien dat het beschikbaar stellen van additionele middelen voor scholen niet leidde tot betere prestaties van leerlingen.

Klassenverkleining

Veel scholen in Nederland gebruiken de compenserende bekostiging voor het verkleinen van klassen (Ladd, 2009). De afgelopen jaren zijn verschillende studies uitgevoerd naar de effecten van klassenverkleining op basis van een experimenteel onderzoeksdesign met aselechte controlegroepen. Het meest bekend is het zogenoemde Star-experiment in Tennessee (Krueger, 1999) waarbij leerlingen en docenten door loting werden toegewezen aan kleine en grote klassen. Daarnaast zijn diverse studies verschenen op basis van natuurlijke experimenten. Zo maken Angrist and Lavy (1999) gebruik van een financieringsregel waardoor de maximale grootte van een klas 40 leerlingen is, en bij 41 leerlingen financiering beschikbaar komt voor een tweede docent (en klas). Deze regel leidt tot toevallige variatie die benut kan worden om het oorzakelijk effect van de klassengrootte vast te stellen. In het algemeen wordt gevonden dat klassenverkleining leidt tot een kleine verbetering van de leerprestaties. De leerwinst is daarbij het hoogst voor leerlingen met een lagere sociaal-economische achtergrond. Recent zijn voor Denemarken (Browning and Heinesen, 2003) en Bolivia (Urquiola, 2006) grotere effecten gevonden. Daar staat tegenover dat voor de Verenigde Staten (Hoxby, 2000) en Noorwegen (Ronning and Oosterbeek, 2006) geen effecten worden gevonden van het verkleinen van klassen. Voor Nederland is in een quasi-experimentele studie, ook gebaseerd op financieringsregels, geen effect gevonden van kleinere klassen gevonden op de leerprestaties (Dobbelsteen, et al. 2002). Klassenverkleining vereist veel middelen en daar staat slechts een geringe verbetering van de leerprestaties tegenover. De empirische literatuur laat zien dat

andere interventies kosteneffectiever zijn, bijvoorbeeld voor- en vroegschoolse educatie of scholing van leraren (zie het voorgaande). Daarnaast heeft het verkleinen van klassen als nadeel dat dit leidt tot een grotere vraag naar leraren. In een periode waarin in Nederland tekorten aan leraren worden verwacht kan dit niet alleen knelpunten opleveren bij de uitvoering van het beleid maar ook de kosten verhogen.

3.6.3 Voortijdig schoolverlaten

Het percentage voortijdig schoolverlaters (VSV-ers) bedroeg in Nederland in 2007 12%.

Hiermee staat Nederland binnen de EU op de 7^e plaats (Lissabon Monitor 2009).

Beleidsintensivering gericht op vermindering van schooluitval zonder startkwalificatie heeft tot doel het bereikte onderwijsniveau van jongeren te verhogen, en daarmee hun latere maatschappelijke positie en kansen op de arbeidsmarkt te versterken. De kansen op de arbeidsmarkt en de sociale integratie verbeteren naarmate men een hoger opleidingsniveau heeft bereikt. Het voorkomen van voortijdig schoolverlaten kan het beroep op sociale uitkeringen verminderen alsmede de kans op sociale uitsluiting en deelname aan criminele activiteiten verkleinen. Hieraan zijn maatschappelijke baten verbonden.

Is dit beleid effectief? Dit hangt af van het instrument dat wordt ingezet. In de praktijk blijken veel programma's niet te werken en blijkt de effectiviteit vaak niet bekend te zijn. In Nederland is in het schooljaar 2006-2007 gestart met het zogenoemde convenantenbeleid. Met dit beleid worden regio's beloond als zij het aantal voortijdig schoolverlaters weten te verminderen. Een evaluatie die gebruik maakt van de gefaseerde invoering van het beleid laat echter zien dat dit beleid geen effect heeft gehad (Van der Steeg et al., 2008). In het schooljaar 2007-2008 is het convenantenbeleid, met een gewijzigde opzet, ingevoerd in heel Nederland. De financiële prikkel is in het nieuwe beleid direct gericht op scholen.

Een inventarisatie van studies met een experimentele opzet (inclusief controlegroepen) geeft twee veelbelovende richtingen (Van der Steeg en Webbink, 2006). In de eerste plaats blijken projecten die gebruik maken van financiële prikkels voor leerlingen (paragraaf 3.6.3.1), leraren en scholen effectief te zijn (zie paragraaf 3.6.4). Positieve ervaringen zijn hiermee opgedaan in het Verenigd Koninkrijk, de VS, en Israël (Cunha, et al. 2005, Angrist and Lavy, 2004, Lavy, 2002, Dearden et al. 2005). Een tweede veelbelovende richting is intensieve coaching, gericht op met name de non-cognitieve ontwikkeling van jongeren met een verhoogd risico op voortijdig schoolverlaten (zie paragraaf 3.6.3.2). Vooral programma's die financiële prikkels voor risicojongeren combineren met het stimuleren van het gebruik van intensieve begeleiding lijken effectief.

Prikkels voor leerlingen

Een beloning voor leerlingen kan mogelijk leiden tot betere onderwijsprestaties. Wanneer leerlingen bijvoorbeeld weinig investeren in human capital omdat ze eerder willen gaan werken, of wanneer zij druk vanuit hun sociale omgeving ervaren om niet te studeren, kan een financiële prikkel ervoor zorgen dat zij meer gemotiveerd raken voor schoolwerk. Uit de psychologische literatuur blijkt echter ook enige evidentie voor mogelijke ongewenste effecten van prikkels. Zo kan het effect van een prikkel bijvoorbeeld sterk afhankelijk zijn de intrinsieke motivatie en van de competenties van een individu (Camerer en Hogarth, 1999). Het effect van prikkels voor studenten blijft daarmee een empirische vraag.

Recent is er bij zowel economen en onderwijsonderzoekers als beleidsmakers een toenemende interesse ontstaan in de rol die prikkels kunnen spelen in het verbeteren van onderwijsprestaties. Deze interesse komt mede voort vanuit het beeld dat het vergroten van onderwijsinputs, zoals meer computers of kleinere klassen, een relatief ineffectieve manier lijkt om onderwijsresultaten te verbeteren (zie o.a. Hanushek, 1986, 1996; Hoxby, 2000). Vanuit het beleid zijn in verschillende landen programma's opgestart met een bepaalde prikkelstructuur voor studenten. Een voorbeeld uit de VS is *No Child Left Behind*, waarin straffen worden opgelegd aan leerlingen die niet aan bepaalde prestatieverplichtingen voldoen. Andere voorbeelden van recent ingevoerde programma's zijn *Education Maintenance Allowance (EMA)* in Engeland en het *PACES*-programma in Colombia. Het *EMA*-programma geeft families met lage inkomens een vergoeding voor deelname aan het onderwijs na de leerplichtige leeftijd en voor het behalen van onderwijsprestaties (Dearden et al., 2001; 2002). *PACES* is een programma waarbij meer dan 125.000 leerlingen vouchers ontvangen die ongeveer de helft van de kosten van het basisonderwijs dekken. De vouchers worden alleen verlengd als leerlingen naar tevredenheid presteren (Angrist et al., 2002). Deze programma's kennen geen experimentele setup, maar geven aan dat prestatieprikkels voor leerlingen een onderwerp is dat ook bij beleidsmakers erg in de belangstelling staat.

Angrist en Lavy (2002) analyseren de effecten van twee beloningsprogramma's voor zwak presterende leerlingen in Israël. Het eerste programma selecteerde een gerandomiseerde groep studenten binnen een school, die een bonus van 800 dollar kon verdienen (of 1000 tot 1200 dollar aan onderwijsvouchers) wanneer het matriculation certificaat in het voortgezet onderwijs behaald werd. In het tweede programma werd eerst een groep van 40 scholen met lage slagingspercentages geselecteerd. Vervolgens werden prestatieprikkels aangeboden aan 20 aselect gekozen scholen uit deze groep. Leerlingen van deze scholen konden bonussen van 500 tot 1500 dollar verdienen. In het eerste programma worden geen significante effecten van financiële prikkels op onderwijsprestaties gevonden. In het tweede programma verhoogt de prikkel de slagingspercentages met 6 tot 8%.

Kremer et al. (2008) analyseren de effecten van beloningen voor meisjes op scholen in Kenia met behulp van een gerandomiseerd experiment. Meisjes op aselect gekozen scholen kregen hun schoolgeld betaald en ontvingen een subsidie voor schoolbenodigdheden wanneer zij goed scoorden op examens. De resultaten suggereren dat deze financiële prikkel tot een significante verbetering van examenresultaten leidt. Zo scoorden meisjes die in aanmerking kwamen voor de beloning gemiddeld 0.12 tot 0.19 standaarddeviatie hoger op de examens. De auteurs vinden bovendien aanwijzingen voor positieve externaliteiten van het beloningsprogramma: ook jongens (die geen kans maakten op de beloning) en meisjes met lage initiële toetsscores (die weinig kans maakten op de beloning) haalden hogere toetsscores.

Bettinger (2008) evalueert de effecten van financiële prikkels in Coshocton, Ohio. Basisscholen in Coshocton bieden sinds 2004 financiële bonussen aan studenten die hun gestandaardiseerde toetsen goed volbrengen. Het programma kent een aselecte toewijzing van leerlingen die in aanmerking komen voor de beloning. Op klas-schoolniveau hielden scholen loterijen aan het begin van de schooljaren 2004-2005, 2005-2006 en 2006-2007. Op deze manier werd in elk schooljaar de helft van de leerlingen in de klassen 3 tot en met 6 geselecteerd voor deelname aan het programma. De beloning bedroeg maximaal 100 dollar per leerling. Het feit dat in opvolgende jaren aparte loterijen worden gehouden, zorgt ervoor dat veel leerlingen het ene jaar wel in aanmerking komen voor de bonus en het andere jaar niet. De auteurs benutten deze manier van randomisering voor het analyseren van de effecten van het programma, waarbij ook de tijdelijke en de 'year-after' effecten geïdentificeerd kunnen worden. De resultaten suggereren dat jonge leerlingen reageren op prikkels. Toetsscores op rekenen waren 0.15 standaarddeviatie hoger bij de leerlingen die in aanmerking kwamen voor de beloning vergeleken met leerlingen in de controlegroep. Dit effect treedt op over de gehele verdeling van toetsscores op rekenen. Voor taal, sociale wetenschap en natuurwetenschappen worden echter minder duidelijke effecten gevonden. Hoewel er positieve tijdelijke effecten gevonden worden, suggereren de resultaten dat deze effecten niet aanhouden in opvolgende jaren wanneer leerlingen niet langer in aanmerking komen voor de bonus.

Uit bovengenoemde studies kan geconcludeerd worden dat financiële prikkels voor leerlingen mogelijk een positief effect hebben op studieprestaties. In enkele (deel)studies worden significante verbeteringen gevonden in studieprestaties. Er worden echter ook insignificante effecten en zelfs één negatief effect op een deelgroep gevonden. Ook is nog niet goed bekend of positieve effecten werkelijk komen door meer inspanning van leerlingen of door andere (mogelijk ongewenste) mechanismen.

Intensieve begeleiding van leerlingen

Ook zijn veelbelovende ervaringen opgedaan met langdurige en intensieve programma's met coaches, gericht op de sociale ontwikkeling van risicjongeren. Heckman (2000) concludeert op basis van bestudering van een groot aantal uitvalbestrijdingsprogramma's dat *langdurige* interventies gericht op jongeren die zich nog steeds in het onderwijs bevinden op het moment van aanvang van de interventie een positieve, doch beperkte invloed kunnen hebben op de studieprestaties en de latere arbeidsmarktpositie. De studies laten volgens Heckman ook zien dat interventies die gericht zijn op jongeren die al zijn uitgevallen, dat wil zeggen curatieve maatregelen, veel minder succesvol zijn. Voorbeelden van effectieve programma's die (mede) gebaseerd zijn op langdurige, intensieve begeleiding van risicoleerlingen zijn Big Brothers/Big Sisters, Sponsor-A-Scholar(SAS), Learning, earning and parenting program (LEAP) en het Quantum Opportunity Program (QOP) in de Verenigde Staten. Wat opvalt in al deze programma's is dat leerlingen een persoonlijke mentor (coach) krijgen toegewezen die zich voornamelijk bezighoudt met de bredere (sociale) ontwikkeling van de jongere, en (veel) minder met de ontwikkeling op school. In het Big Brothers/Big Sisters programma werden zelfs helemaal geen specifieke pogingen door de mentor ondernomen om schoolachterstanden in te halen of bepaalde onderwijsdoelen te halen. De mentor heeft in dit programma meer een brede, ondersteunende rol en fungeert als een vriend en rolmodel. Toch zijn positieve resultaten op schoolprestaties gevonden. Het quantum opportunity program combineert een groepsmentor met financiële prikkels voor high school leerlingen uit lagere sociaal-economische milieus. De toewijzing van studenten aan het programma gebeurde random. De financiële prikkels houden een beloning in van twee US \$ voor elk uur besteed aan activiteiten gericht op verbetering van sociale vaardigheden of persoonlijke ontwikkeling. Naast positieve effecten op schoolprestaties en een langer verblijf in het onderwijs worden ook sociale baten gevonden in de vorm van significant minder crimineel gedrag.

Dynarski et al. (1998) hebben 20 Amerikaanse uitvalbestrijdingsprogramma's geëvalueerd die tussen 1991 en 1996 zijn uitgevoerd onder het *School Dropout Demonstration Assistance Program* (SDDAP). Zestien van deze programma's hadden een experimentele opzet, waardoor causale effecten konden worden vastgesteld. Het betreft zowel projecten die gericht zijn op een herstructurering van scholen - zogenoemde 'alternative schools' - als projecten die minder intensieve diensten (bijvoorbeeld tutor programma's of incidentele lessen gericht op het vergroten van het gevoel van eigenwaarde) bieden aan achterstandsleerlingen in de bestaande settings - zogenoemde 'supplemental programs'. De helft van de programma's is gericht op 'middle schools', de andere helft op 'high schools'.

De conclusie van de totale evaluatie is dat de meeste programma's niet of nauwelijks hebben geleid tot een vermindering van de schooluitval of een verbetering van de schoolprestaties. Als oorzaak voor deze geringe impact wordt genoemd een gebrekkige kennis van de diverse oorzaken van vsy, alsmede de grote variatie aan problemen van leerlingen die

een gedifferentieerde aanpak van vsv noodzakelijk maakt (in tegenstelling tot de 'one size fits all'-benadering die vaak wordt toegepast).

Voor enkele projecten zijn echter wel positieve resultaten gevonden. Op 'middle school'-niveau gaat het dan om de 'alternative middle schools'-programma's, die worden gekenmerkt door een intensieve en persoonlijke benadering van risicoleerlingen in kleine groepen. Deze programma's droegen echter alleen bij aan een lagere uitval, en niet aan betere studieprestaties. In drie van de vier programma's was bovendien zelfs sprake van een hoger schoolverzuim onder de deelnemers dan onder de jongeren in de controlegroep.

Op 'high school'-niveau zijn enkele positieve effecten gevonden van *GED-programma's* op het behalen van een diploma. Deze programma's bieden jongeren die (meestal) al zijn uitgevallen hulp om alsnog een certificaat te halen, het zogenoemde 'GED-certificate' (General Educational Development). Er is echter geen sprake van een afname van de dropout rate. In recent onderzoek worden bovendien twijfels geplaatst bij de economische waarde van een GED-certificaat, in termen van de invloed op het salaris en de kans op werk ten opzichte van dropouts. Een andere kanttekening is dat een groot gedeelte (circa tweederde) van de deelnemers aan de GED-programma's het programma uiteindelijk niet afmaakt.

De effectiviteit van de herstructureringsprogramma's van scholen blijkt in algemene zin sterk bepaald te worden door de mate waarin 'commitment' aanwezig is bij de scholen (directeuren en leraren) om wijzigingen door te voeren. Dynarski et al. (1998) concluderen dat herstructureringsprogramma's alleen effectief lijken te zijn als er werkelijk iets veranderd wordt aan wat er in de klas zelf gebeurt. Tymms ea. (2009) hebben recent een grootschalig experiment uitgevoerd waarin ze leerlingen random toewezen aan verschillende vormen van 'peer tutoring'. Uit deze studie blijkt dat 'peer tutoring' de taal- en rekenprestaties van basisschoolleerlingen verbetert. Door te variëren in condities tonen de auteurs ook aan welke vormen en intensiteit van begeleiding het meest effectief is.

Een neveneffect van dit beleid met langdurige en intensieve programma's is dat het de toekomstige inkomensverschillen vermindert. Wegen de baten op tegen de kosten? Bovengenoemd economisch onderzoek leert dat de baten kunnen opwegen tegen de kosten. Echter, van veel projecten is de effectiviteit niet bekend. Het is daarom belangrijk hierover meer kennis te verwerven.

3.7 Excellentie

Een vergelijking van de prestaties van Nederlandse leerlingen met leerlingen uit andere rijke landen laat zien dat Nederland goed scoort aan de linkerkant van de verdeling van kennis en vaardigheden, maar minder goed scoort aan de rechterkant van de verdeling (Minne et al. 2007). Anders gezegd, qua excellentie lijkt er ruimte voor verbetering te zijn voor het Nederlandse onderwijssysteem. Recent onderzoek levert aanwijzingen dat excellentie in het onderwijs de welvaart bevordert zowel voor het individu als voor de samenleving als geheel.

Over het bevorderen van excellentie is echter weinig bekend. De empirische literatuur levert weinig inzicht in welk beleid hiervoor het meest effectief is. Financiële prikkels voor studenten blijken de prestaties van de betere leerlingen te bevorderen (maar geen effect te hebben op de overige leerlingen). Een toename van private middelen voor het hoger onderwijs biedt de mogelijkheid om de kwaliteit van opleidingen te verhogen. De invloed van het verhogen van private bijdragen op de toegankelijkheid van het hoger onderwijs, lijkt op basis van Amerikaanse studies, klein te zijn. In een recente Amerikaanse studie zijn echter ook aanwijzingen voor leenangst gevonden.

Voor het Nederlandse onderwijs lijkt de vraag aan de orde te zijn op welke wijze de excellentie het best kan worden bevorderd. Een eerste vraag is waar nieuw beleid het best bij kan aangrijpen: bij hoger onderwijs, voorgezet onderwijs of meteen al bij het basisonderwijs? Nader onderzoek kan zich ook richten op het effect van recent geïntroduceerde programma's voor excellente leerlingen in het voortgezet en hoger onderwijs. Ook zijn er vragen rond de inrichting (selectie aan de poort) en financiering van het hoger onderwijs (differentiatie van collegegelden).

Het belang van excellentie

Wat zijn de opbrengsten van het volgen van een kwalitatief goede opleiding? Voor de Verenigde Staten is dit onderzocht door de financiële opbrengsten te analyseren van verschillende opleidingen in het hoger onderwijs en daarbij te corrigeren voor verschillen in geschiktheid van deelnemers (Hoxby, 2001). Zij vindt dat het afstuderen aan selectieve instellingen hoge financiële rendementen geeft voor alle opleidingen. De investeringen in opleidingen worden meerdere malen terugverdiend gedurende de beroepsloopbaan. Als afgestudeerden met dezelfde geschiktheid worden vergeleken blijken degenen die een selectieve opleiding hebben gevolgd het meest te verdienen. Recent onderzoek geeft ook inzicht in de sociale opbrengsten van excellentie. Hanushek & Woessmann (2007) vinden sterke aanwijzingen dat het aandeel mensen in de top van verdeling van kennis en vaardigheden, gemeten als het aandeel leerlingen dat behoort tot de besten van de wereld, belangrijk is voor de productiviteit van landen en voor economische groei. Er zijn ook aanwijzingen dat een toename van het aandeel hoger opgeleiden ook bijdraagt aan de productiviteit van anderen (Iranzo & Peri, 2006). De hoge niveaus van kennis en vaardigheden blijken vooral belangrijk te zijn in hoogproductieve landen, zoals Nederland. Naarmate landen dichterbij de zogenoemde 'technology frontier' opereren valt er minder te imiteren van andere landen en wordt innovatie belangrijker voor de groei van de productiviteit (Vandenbussche et al. (2006, Aghion et al. (2005). Zie ook Minne et al. (2007) voor een overzicht van deze literatuur.

Het bevorderen van excellentie

Over het bevorderen van excellentie is weinig bekend. Mogelijkheden daarvoor zijn excellente programma's in verschillende fasen van het onderwijs, financiële prikkels of hogere private bijdragen voor studenten. De empirische literatuur geeft geen inzicht in het effect van het gebruik van excellente programma's in het onderwijs.

Prikkels voor studenten

Leuven et al. (2009) evalueren een gerandomiseerd experiment met prikkels voor studenten in Amsterdam. Eerstejaars studenten economie en business konden een financiële beloning verdienen als zij binnen een jaar slaagden voor alle verplichte eerstejaarsvakken. De studenten werden aselekt verdeeld over drie groepen. Leerlingen in de eerste groep konden een beloning van 681 euro tegemoet zien, leerlingen in de tweede groep kregen een beloning van 227 euro en leerlingen in de derde (controle)groep kwamen niet in aanmerking voor een beloning. Op de hele steekproef van leerlingen worden kleine, niet significante positieve effecten gevonden van een beloning op het slagingspercentage en het aantal behaalde studiepunten. Er worden echter wel significante effecten gevonden voor de betere leerlingen: leerlingen met hoge cijfers voor wiskunde in het voortgezet onderwijs haalden significant meer studiepunten wanneer zij in een (hogere) beloningsgroep werden geselecteerd. Aan de andere kant bleken zwakkere leerlingen juist slechter te presteren wanneer zij in een hogere beloningsgroep zaten. De auteurs beargumenteren dan ook dat het totale effect van de prestatiebeloning niet insignificant is omdat het een gemiddelde is van een positief effect aan de bovenkant van de verdeling en een negatief effect aan de onderkant.

Angrist et al. (2006) evalueren de effecten van een gecontroleerd veldexperiment op een grote Canadese universiteit dat als doel heeft de prestaties van eerstejaarsstudenten te verbeteren. Een groep studenten kreeg aselekt prestatie-afhankelijke beloningen aangeboden voor goede (maar niet noodzakelijk top-) cijfers in het eerste jaar. Het programma had significant positieve effecten op de cijfers van vrouwelijke studenten in het eerste blok. De effecten ebden iets weg op het einde van het eerste jaar, maar bleven significant voor vrouwen die genoeg vakken wilden halen om zich te kwalificeren voor een studiebeurs. Ook wordt een verlagend effect gevonden op uitval. Dit effect is groot genoeg om een overall verlaging van uitval te bewerkstelligen. Er zijn geen effecten gevonden op de prestaties van mannelijke studenten. De resultaten waren het meest veelbelovend voor de groep studenten die naast de financiële prikkels ook adviezen door peers en georganiseerde groepstudie kregen aangeboden.

Private bijdragen hoger onderwijs

Hogere private bijdragen verschaffen meer middelen aan instellingen waarmee de kwaliteit van opleidingen kan worden verhoogd. In de empirische literatuur is vooral gekeken naar het effect van hogere private bijdragen op de deelname aan het hoger onderwijs. Het algemene beeld uit de literatuur is dat de prijselasticiteit van de vraag naar hoger onderwijs laag is (Jacobs, 2002).

In een groot aantal studies in de Verenigde Staten is gevonden dat een verlaging van de collegegelden met 1000 dollar leidt tot 3 tot 4 procent meer deelname aan hoger onderwijs (Dynarski, 2003). De lage prijselasticiteit heeft waarschijnlijk te maken met de hoge opbrengsten van studeren in het hoger onderwijs. De loonverschillen tussen hoger opgeleiden en middelbaar opgeleiden zijn aanzienlijk. Gemiddeld verdienen Hbo'ers per jaar hoger onderwijs 7 % meer en Wo'ers 9 % meer. Recent onderzoek suggereert dat dit loonverschil daadwerkelijk veroorzaakt wordt door het extra onderwijs en niet door andere factoren (Webbink, 2007). In Australië is in 1989 een sociaal leenstelsel geïntroduceerd in de vorm van het zogenoemde Higher Education Contribution Scheme (HECS). Voor de introductie van het HECS kende Australië geen private bijdragen aan hoger onderwijs. Door de introductie steeg de private bijdrage in de gemiddelde kosten van hoger onderwijsprogramma's naar 23%. Verschillende evaluaties laten zien dat de deelname aan het hoger onderwijs in Australië niet is afgenomen na de introductie van het HECS (zie voor een overzicht CPB, 2003). Dit is in lijn met de resultaten van empirisch onderzoek naar de prijselasticiteit van de vraag naar hoger onderwijs: wat is het effect van hogere collegegelden op de deelname aan hoger onderwijs? In een recente studie in de Verenigde Staten (Field, 2009) is leenangst onderzocht: aan het hebben van een schuld zijn mogelijk 'psychologische kosten' verbonden. In een experiment werd aan rechtenstudenten twee soorten studiefinanciering aangeboden die monetair equivalent waren. De toekenning gebeurde door middel van loting. Voor beide soorten studiefinanciering gold dat de schuld volledig werd kwijtgescholden als men tien jaar bij de overheid ging werken. In het eerste studiefinancieringspakket werd ook een beurs toegekend, die weer omgezet werd in een lening als men niet bij de overheid ging werken. In het tweede pakket werd alleen een lening aangeboden. Gevonden werd dat het type studiefinanciering van invloed was op de onderwijsdeelname en op de keuze van het beroep na de studie. Als leenaversie daadwerkelijk belangrijk is voor de deelname aan hoger onderwijs kan dit een reden zijn in de studiefinanciering ook subsidie-elementen in te bouwen, bijvoorbeeld voor kwetsbare groepen. De kennis over leenaversie bij de deelname aan hoger onderwijs in Nederland is echter nog beperkt.

3.8 School en omgeving

De laatste jaren worden scholen steeds meer verbonden met andere actoren of instellingen uit de omgeving van de school. Een voorbeeld hiervan is de zogenoemde 'Brede school', waarin de school bijvoorbeeld samenwerkt met culturele en sportinstellingen en er ook na de reguliere schooltijd activiteiten zijn voor leerlingen of kinderen uit de buurt. Onderzoek naar deze activiteiten suggereert dat leerlingen op deze scholen beter presteren. Zo vindt Sheldon (2007, 2005) bijvoorbeeld dat de aanwezigheid van leerlingen hoger is op scholen die programma's ontwikkelen waarbij school, familie en gemeenschap betrokken zijn en dat specifieke betrokkenheid van de familie of de omgeving een positief effect heeft op wiskunde toetscores

van leerlingen. Het gaat hierbij echter niet om studies met een experimentele opzet. Het is daarom niet duidelijk of de 'brede school' daadwerkelijk leidt tot positieve uitkomsten omdat de uitkomsten vertekend kunnen zijn door de selectie van ouders en scholen. Een opvallende ontwikkeling in de omgeving van het Nederlandse publieke onderwijs is de groei van huiswerkbegeleiding door private aanbieders. Er bestaan geen experimentele studies naar de effecten hiervan op de prestaties van leerlingen.

Voor het Nederlandse onderwijsstelsel lijken daarom vragen aan de orde over de effecten van 'brede scholen' en van huiswerkbegeleiding door private aanbieders.

4 Onderzoeksagenda

Als we het Nederlandse onderwijsstelsel afzetten tegen de empirische evidentie blijkt dat de opzet van het Nederlandse stelsel op verschillende punten als, in potentie, effectief kan worden gekarakteriseerd. Nederland kent al geruime tijd een vrije schoolkeuze en concurrentie tussen scholen. Nederlandse scholen hebben een grote mate van autonomie. Centrale eindexamens zorgen voor transparantie over de prestaties van leerlingen en scholen. Daarnaast bewaakt de Onderwijsinspectie de kwaliteit van het onderwijs. Ook volgt de aanpak van achterstanden in het onderwijs op hoofdlijnen de meest kansrijke benadering volgens de empirische evidentie. Voor- en vroegschoolse programma's worden steeds meer ingezet en veel achterstandsmiddelen worden ingezet bij de jongste leerlingen. Het is echter niet duidelijk of deze, in potentie, sterke elementen van het Nederlandse onderwijs, volledig tot hun recht komen en daadwerkelijk hun doel bereiken. Hierbij speelt de vormgeving van beleid een belangrijke rol. Vaak is niet duidelijk of de gekozen aanpakken en interventies daadwerkelijk effectief zijn.

Transparantie informatie over de kwaliteit van scholen

Vanuit internationaal perspectief heeft het Nederlandse onderwijsstelsel een aantal opvallende kenmerken: vrije schoolkeuze, concurrentie tussen scholen en een grote mate van autonomie voor scholen. Voor een goede werking van een dergelijk stelsel is transparante informatie over de kwaliteit (prestaties) van scholen van groot belang. Hiervoor zijn twee redenen. Allereerst stelt een goede informatievoorziening ouders in staat om een goede school te kiezen. In de tweede plaats is transparantie informatie over prestaties van scholen belangrijk als tegenwicht voor de autonomie van scholen.

Een vrije schoolkeuze is gebaat bij een goede informatievoorziening over de kwaliteit van scholen en bij een goed gebruik van deze informatie bij de schoolkeuze. Wanneer slechts een deel van de ouders de beschikbare informatie gebruikt zal vrije schoolkeuze maar beperkt bijdragen aan de concurrentie. Als dit gebruik van informatie sterk varieert naar sociaaleconomisch achtergrond kan vrije schoolkeuze bovendien zorgen voor meer ongelijkheid naar sociaaleconomische achtergrond. De recente plannen in de gemeente Nijmegen, om via een centraal meldpunt te komen tot een andere spreiding van leerlingen over scholen, lijken ook ingegeven door bezorgdheid over de uitkomsten van de vrije schoolkeuze. Daarom lijkt de vraag aan de orde in hoeverre de schoolkeuze in Nederland leidt tot gewenste resultaten, en in hoeverre de informatievoorziening daaraan bijdraagt of aanpassing behoeft. Recent onderzoek in de Verenigde Staten laat zien dat eenvoudig toegankelijke en transparante informatie over de kwaliteit van scholen invloed heeft op het keuzegedrag van ouders en leerlingen, en dat overstappers beter presteren op de nieuw gekozen scholen. Onderzoek naar de rol van informatie bij de schoolkeuze in Nederland lijkt daarom gewenst.

Het Nederlandse onderwijs kent een grote mate van autonomie van scholen. Transparante informatie over de kwaliteit van scholen werkt disciplinerend en stimuleert scholen tot goede

prestaties. Het ontwerp van goede beoordelingssystemen van scholen is daarom belangrijk voor Nederland. Daarnaast spelen vragen over hoe om te gaan met slecht presterende scholen. Falende scholen lijken de achilleshiel van het Nederlandse stelsel. Over de omgang met falende scholen bestaat nauwelijks harde evidentie. Wel zijn er aanwijzingen dat zwak presterende scholen in staat zijn zich te verbeteren. Hoe dit het best gerealiseerd kan worden, door ondersteuning of door prikkels en sancties, is niet duidelijk. Onderzoek naar beste aanpak voor falende of slecht presterende scholen lijkt daarom belangrijk voor het Nederlandse onderwijs.

De kwaliteit van leraren en schoolleiders

Leraren spelen een sleutelrol in het onderwijs. Een cruciale vraag betreft het bevorderen van de kwaliteit van leraren. In het bijzonder spelen er vragen over de aansluiting tussen opleiding en beroep. De empirische literatuur vindt geen sterke relatie tussen enerzijds de initiële opleiding of de toetsscores van de docent en anderzijds de prestaties van leerlingen. Wat betekent dit voor de werving van kwalitatief goede leraren? Zijn opleidingseisen nodig of vormen ze juist een obstakel voor het aantrekken van goed personeel doordat ze het potentiële aanbod verminderen? Een andere belangrijke vraag betreft de beloning van leraren. De beloning van leraren en schoolleiders is vooral gebaseerd op ervaring. De empirische literatuur leert dat ervaring nauwelijks van invloed is op de prestaties van leraren. Alleen in de eerste twee jaren heeft het ontbreken van ervaring een negatief effect. Recente studies vinden positieve effecten van het gebruik van prestatiebeloning in het onderwijs. Dit roept de vraag op naar de mogelijkheden om de beloning van leraren en schoolleiders te relateren aan hun prestaties. Wat zijn de mogelijkheden voor en effecten van prestatiebeloning in het Nederlandse onderwijs?

Centrale examens versus schoolexamens

De empirische literatuur geeft stevig bewijs voor positieve effecten van het gebruik van centrale examens op de kwaliteit van het onderwijs. Dit roept de vraag op of het gebruik van schoolexamens, naast centrale examens, nadelig is voor de kwaliteit van het onderwijs?

Vroege selectie

Een opvallend aspect aan het Nederlandse onderwijs is de vroege selectie van leerlingen na de basisschool. Recent onderzoek suggereert dat vroege selectie voor bepaalde groepen leerlingen de kansen vermindert op het behalen van een hoger onderwijs diploma. Een belangrijke vraag is op welke wijze de eventuele nadelige effecten van vroege selectie van bepaalde groepen leerlingen kunnen worden tegengegaan, hetzij door uitstel van selectie van leerlingen, hetzij door het vergroten van latere doorstroommogelijkheden.

Effectieve interventies

Op verschillende terreinen zijn vragen aan de orde over de effectiviteit van de gekozen aanpakken en interventies. Zo is over de effectiviteit van de stelselkenmerken van het speciaal onderwijs weinig bekend. Nederland kent de laatste jaren een verschuiving van leerlingen naar duurere vormen van speciaal onderwijs. Wat zijn de effecten van deze vormen van speciaal onderwijs voor de betreffende leerlingen maar ook voor klasgenoten bij opvang in het regulier onderwijs? Bij de aanpak van onderwijsachterstanden in Nederland is het de vraag welke interventies het meest kosteneffectief zijn, en hoe scholen kunnen worden gestimuleerd om deze aanpakken ook daadwerkelijk te gebruiken? Excellentie in het onderwijs is een relatief nieuw onderwerp op de beleidsagenda. Ook hier spelen vragen over de effectiviteit van beleid gericht op het bevorderen van excellentie. Waar kan beleid het best aangrijpen: bij hoger onderwijs, voorgezet onderwijs of meteen al bij het basisonderwijs? Nader onderzoek kan zich ook richten op het effect van recent geïntroduceerde programma's voor excellente leerlingen in het voortgezet en hoger onderwijs. Tot slot, is over de effecten van de brede school of van huiswerkbegeleiding door private instellingen weinig bekend.

Welk type onderzoek?

Het vaststellen van de effecten van kenmerken van het onderwijsstelsel is lastig omdat in het onderwijs tal van selectieprocessen een rol spelen. Ouders en leerlingen kiezen scholen, daarnaast maken scholen voortdurend keuzen met betrekking tot het leerproces en de inzet van docenten. Veranderingen in het onderwijsstelsel kunnen grote effecten hebben en brengen derhalve risico's met zich. Deze risico's kunnen worden verkleind door beleidsbeslissingen te onderbouwen met kwalitatief hoogwaardig onderzoek. Dit betreft onderzoek waarin rekening wordt gehouden met selectiviteit door gebruik te maken van geloofwaardige onderzoeksdesigns, zoals experimenten en natuurlijke experimenten.

Opvallend aan deze overzichtsstudie is dat deze weinig Nederlandse studies met een (quasi)experimenteel design bevat. Het werken met controlegroepen en voor- en nametingen is nog relatief nieuw in het Nederlandse onderwijsonderzoek, en random toewijzing is een uitzondering. Vanuit de onderzoeks-, beleids- en onderwijswereld worden vaak diverse redenen aangedragen om niet de hoogste methodologische eisen te stellen. Een vergelijkbare situatie bestond enkele jaren geleden in Amerika.¹⁵ Inmiddels is dit type onderzoek daar wel doorgebroken. In Nederland gaan binnenkort ook experimenten van start ondersteund door middelen vanuit het actieprogramma Onderwijsbewijs.¹⁶ Het bevorderen van dit type onderzoek, dat solide bewijs oplevert voor beleidsbeslissingen, kan hoge maatschappelijke opbrengsten genereren. De kosten van slecht onderbouwd beleid, dat na verloop van tijd niet blijkt te werken, zijn namelijk zeer hoog.

¹⁵ Zie Cook (2002) voor een bespreking van discussie over experimenteel onderzoek in het onderwijs.

¹⁶ <http://www.minocw.nl/onderwijsbewijs/index.html>

Lerend beleid

Nieuw beleid is de afgelopen decennia in het Nederlandse onderwijs vaak over de hele linie ingevoerd. Als dit beleid niet werkt kan de schade groot zijn. Bovendien is het lastig om de effecten vast te stellen omdat er geen controlegroepen zijn als beleid over de hele linie wordt ingevoerd. Anders gezegd, er zijn grote risico's verbonden aan het invoeren van nieuw beleid over de gehele linie. Deze risico's kunnen worden verminderd door nieuw beleid eerst op kleinere schaal in te voeren. Als daarbij gekozen wordt voor een opzet met een experimenteel design kan solide bewijs worden verkregen over de effecten van het nieuwe beleid. Dit bewijs kan vervolgens worden gebruikt bij de beslissing om het nieuwe beleid geheel in te voeren, dan wel bij te stellen of stop te zetten. Mogelijkheden voor dit lerend beleid doen zich voor bij beleid dat gefaseerd wordt ingevoerd, zoals de eerste ronde van het convenantenbeleid ten aanzien van het voortijdig schoolverlaten. Een andere mogelijkheid is om bij pilotprojecten gebruik te maken van controlegroepen. Vaak zijn er onvoldoende middelen om alle leerlingen aan een nieuwe interventie te kunnen laten deelnemen. In dat soort situaties kan door loting worden bepaald wie wel en wie niet aan de nieuwe interventie gaan deelnemen. Door te loten heeft iedereen evenveel kans. Het benutten van dit soort mogelijkheden vereist een goede afstemming tussen beleid en onderzoek. De meeste onderdelen van de hierboven geschetste onderzoeksagenda lenen zich voor een dergelijke aanpak.

Referenties

Concurrentie, vrije schoolkeuze en segregatie

Ammermueller, A. and J.S. Pischke, 2006, Peer effects in European primary schools, evidence from PIRLS, NBER working paper 12180.

Angrist, J.D., K. Lang, 2004, Does school integration generate peer effects? Evidence from Boston's Metco Program, *American Economic Review*, 94 (5), 1613-1634.

Belfield, C. R. and H.M. Levin, 2003, The Effects of Competition on Educational Outcomes: A Review for the United States, Occasional Paper 35, Columbia University, New York, NY. National Center for the Study of Privatization in Education.

Corten, R. and J. Dronkers, 2006, School achievement of pupils from the lower strata in public, private government-dependent and private dependent-independent schools: a cross-national test of the Coleman-Hofer thesis, *Educational Research and evaluation*, 12, 179-208.

Cullen, J.B, B.A. Jacob, S. Levitt, 2006, The effect of school choice on participants: Evidence from randomized lotteries, *Econometrica* 74, 1191-1230.

Cullen, J.B, B.A. Jacob, 2007, Is gaining access to selective elementary schools gaining ground? Evidence from randomized lotteries?, NBER working paper 13443.

Ding, W. and S.F. Lehrer, 2007, Do peers affect student achievement in China's secondary schools, *Review of Economics and Statistics*, 89(2), 300-312.

Driessen, G., 2007, Peer group effecten op onderwijsprestaties, Een international review van effecten, verklaringen en theoretische en methodologische aspecten, ITS, Radboud Universiteit Nijmegen.

Dronkers, J. and P. Robert, 2008, Differences in scholastic achievement of public, private government-dependent, and private dependent schools: A cross-national analysis, *Educational Policy*, 22, in press.

Fuchs, T. and L. Woessmann, 2007, What accounts for international differences in primary schools learning across countries, Mimeo, IFO Institute for Economic Research at the University of Munich.

- Gibbons, S., S. Machin and O. Silva, 2008, Choice, Competition and Pupil Achievement, *Journal of the European Economic Association*, 6(4), pp. 912-947.
- Gould, E.D., Lavy, V. , M.D. Paserman, 2004, Does immigration affect the long-term outcomes of natives, quasi-experimental evidence, NBER Working Paper 10844.
- Guryan, J., 2004, Desegregation and black dropout rates, *American Economic Review*, 94 (4), 919-941.
- Hastings, J., J.M. Weinstein, 2008, Information, school choice and academic achievements: evidence from two experiments, *Quarterly Journal of Economics*, 123 (4), 1373-1414.
- Holmes, G. M., J. DeSimone and N.G. Rupp, 2003, Does School Choice Increase School Quality?, NBER Working Paper, WP 9683.
- Hoxby, C., 2000, Peer effects in the classroom: learning from gender and race variation, NBER Working Paper 7867.
- Hoxby, Caroline M., 2000, Does Competition among Public Schools Benefit Students and Taxpayers?, *American Economic Review*, 90(5), pp. 1209-1238.
- Hoxby, C.M. and G. Weingarth-Salyer, 2005, Taking race out of the equation: school reassignment and the structure of peer effects, mimeo, Stanford University.
- Hsieh, C. T., and M. Urquiola, 2003, When Schools Compete, How Do They Compete?, An Assessment of Chile's Nationwide School Voucher Programme, NBER Working Paper, WP 10008.
- Janssens, F. and A.Visscher, 2004, Naar een kwaliteitskaart voor het primair onderwijs, *Pedagogische Studiën*, 81, 5, 371-383.
- Karsten, S., A. Visscher and T. de Jong, 2001, Another side to the coin: The unintended effects of the publication of school performance data in England and France, *Comparative Education*, 37(2), 231-242.
- Kling, J. R., J. B. Liebman, and L. F. Katz, 2007, Experimental Analysis of Neighborhood Effects, *Econometrica*, 75(1), 83-119.

- Krueger and Zhu, 2004, Another look at the New York City school voucher experiment, *American Behavioral Scientist*, Vol. 47, No. 5, 658-698.
- Lavy, V. and A. Schlosser, 2007, Mechanisms and impacts of gender peer effects at school, NBER Working Paper 13292.
- Lefgren, L., 2004, Educational peer effects and the Chicago public schools, *Journal of Urban Economics*, 56(2), 169-191.
- Neal, D., 1997, The effects of Catholic secondary schooling on educational attainment, *Journal of Labor Economics* 15: 98-123.
- Neidell, M. and J. Waldfogel, 2008, Cognitive and non-cognitive peer effects in early education, NBER Working Paper 14277.
- Noailly, J., S. Vujic and A. Aouragh, 2009, The effects of competition on the quality of primary schools in the Netherlands, CPB discussion paper 120.
- Peterson, Howell, Wolf and Campbell, 2003, School vouchers results from randomised experiments. In C.Hoxby (ed.), *The Economic Analysis of School Choice*, University of Chicago Press.
- Sacerdote, B., 2000, Peer effects with random assignment: Results for Dartmouth Roommates, in: *Quarterly Journal of Economics*, vol. 116, 2, mei 2001, p. 681-704.
- Woessman, L., 2001, Why students in some countries do better: international evidence on the importance of education policy, *Education Matters*, 1(2), 67-74.
- Woessman, L., 2003, Schooling resources, educational institutions and student performance: The international evidence, *Oxford Bulletin of Economics and Statistics*, 65 (2), 117-170.
- Woessman, L., 2008, Fundamental Determinants of school efficiency and equity: German States as a microcosm for OECD countries, Harvard University, Program on education policy and governance research paper PEPG 07-02, Cambridge, MA: Harvard University.
- Woessman, L. E. Ludemann, G. Schutz and M.R. West, 2009, *School accountability, autonomy and choice around the World*, Cheltenham: Edward Elgar, in press.

Autonomie en verantwoording

Ayres, I. en J. Braithwaite, 1992, *Responsive Regulation: Transcending the Deregulation Debate*, New York: Oxford University Press, 1992.

Betts, J.R., 1998, The Impact of Educational Standards on the Level and Distribution of Earnings, *American Economic Review*, 88, 266-275.

Bishop, J.H., 1997, The Effect of National Standards and Curriculum-Based Examinations on Achievement, *American Economic Review*, 87, 260-264.

Bishop, J.H., 2006, Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems and Peer Pressure, In: Hanushek, Eric A. and Welch, F. (eds.), *Handbook of the Economics of Education*, Vol. 2, pp. 909-944, Amsterdam: North-Holland.

Black, P. en D. William, 1998, Inside the black box: Raising standards through classroom assessment, *Phi Delta Kappan*, 80, 139-148.

Costrell, R.M., 1994, A Simple Model of Educational Standards, *American Economic Review*, 84, 956-971.

Fuchs, T. and L. Wößmann, 2007, What Accounts for International Differences in Student Performance? A Re-examination using PISA Data, *Empirical Economics*, 32, 433-464.

Fuchs, T. and L. Woessman, 2005, The determinants of differences in primary school learning across countries, mimeo, IFO Institute for Economic Research at the University of Munich.

Hattie, J. and H. Timperley, 2007, The power of feedback, In: *Review of Educational Research*, Vol. 77, No. 1, 81-112.

Hoxby, C.M., 2002, The Cost of Accountability,. In: Evers, W.M., & Walberg, H.J. (eds.), *School Accountability*. Stanford, CA: Hoover Institution Press.

Janssens, F.J.G. en I.F. de Wolf, 2007, Effectmeting voor de Inspectie van het Onderwijs. Keynote-lezing op de Onderwijsresearchdagen, juni 2007.

Jürges, H., K. Schneider and F. Büchel, 2005, The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany, *Journal of the European Economic Association*, 3, 1134-1155.

Kluger, A.N. en A. DeNisi, 1996, The Effects of Feedback Interventions on Performance: Historical Review, a Meta-Analysis and a Preliminary Feedback, In: *Psychological Bulletin*, 119, 254-284.

Wolf, I.F. de en F.J.G. Janssens, 2007, Effects and side effects of school inspections and accountability in education: a review of empirical studies, In: *Oxford Review of Education*, 33, 3, 379-396.

Wößmann, L., 2003a, Schooling Resources, Educational Institutions and Student Performance: the International Evidence, *Oxford Bulletin of Economics and Statistics*, 65, 117-170.

Wößmann, L., 2003b, Central Exams as the “Currency” of School Systems: International Evidence on the Complementarity of School Autonomy and Central Exams, DICE Report – *Journal for Institutional Comparisons*, 1, 46-56.

Selectie

Ammermüller, A., 2005, Educational Opportunities and the Role of Institutions. ZEW Discussion Paper 05-44, Mannheim: Centre for European Economic Research.

Ariga, K. and G. Brunello, 2007, Does secondary school tracking affect performance? Evidence from IALS, IZA Discussion Paper 2643.

Argys, L., D. Rees, and D. Brewer, 1996, Detracking America’s schools: equity at zero cost? *Journal of Policy Analysis and Management*, vol. 15, no. 4, pp. 623-645.

Brunello, G., D. Checchi, 2007, Does tracking affect equality of opportunity? New International Evidence, *Economic Policy*, vol. 52 , pp. 781-861.

Duflo, E., P. Dupas and M. Kremer, 2008, Peer effects and the impact of tracking: Evidence from a Randomized Evaluation in Kenya, NBER Working Paper 14475.

Van Elk, R., M. van der Steeg and D. Webbink, 2009, The effect of early tracking on participation in higher education, The Hague: CPB Netherlands Bureau for Economic Policy Analysis.

Figlio, D. and M. Page, 2002, School choice and the distributional effects of ability tracking: does separation increase inequality?, *Journal of Urban Economics*, vol. 83, pp. 497-514.

Hanushek, E. and L. Woessman, 2006, Does education tracking affect performance and inequality? Differences in differences evidence across countries, *The Economic Journal*, vol. 116, C63-C76.

Schütz, G., H.W. Ursprung and L. Woessmann, 2008, Education Policy and Equality of Opportunity, *Kyklos* 61 (2), S. 279-308.

Leraren

Aaronson, D., L. Barrow and W. Sander, 2007, Teachers and student achievement in the Chicago public high schools, *Journal of Labor Economics*, 25(1): 95-135.

Angrist, J.A. and Victor Lavy, 2001, Does Teacher Training Affect Pupil Learning? Evidence From Matched Comparisons In Jerusalem Public Schools, *Journal of Labor Economics*, 2001, v19 (2, Apr), 343-369.

Atkinson, A., S. Burgess, B. Croxson, P. Gregg, C. Propper, H. Slater, D. Wilson, 2004, Evaluating the Impact of Performance-related Pay for Teachers in England, CMPO Working Paper 04/113, Bristol: Centre for Market and Public Organisation.

Boyd, D., P. Grossman, H. Lankford, S. Loeb and J. Wyckoff, 2005, How Reduced Barriers to Entry into Teaching Changes the Teacher Workforce and Affects Student Achievement, NBER Working Paper 11844.

Cantrell, S., J. Fullerton, T. Kane and D. Staiger, 2008, National board certification and teacher effectiveness: evidence from a random assignment experiment, NBER Working Paper 14608.

Clotfelter, C., H. Ladd and J. Vigdor, 2007, Teacher credentials and student achievement in high school: a cross-subject analysis with student fixed effects, NBER Working Paper 13617.

Eberts, R., K. Hollenbeck and J. Stone, 2002, Teacher performance incentives and student outcomes, *Journal of Human Resources*, 37(4), 913-927.

Figlio, D.N. and L.S. Getzler, 2002, Accountability, ability and disability: Gaming the system, NBER Working Paper 9307.

- Fuchs, T. and L. Wößmann, 2007, What Accounts for International Differences in Student Performance? A Re-examination using PISA Data, *Empirical Economics*, 32, 433-464.
- Glewwe, P., N. Ilias and M. Kremer, 2003, Teacher incentives, NBER Working Paper 9671.
- Gordon, R., T. Kane and D. Staiger, 2006, The Hamilton Project: Identifying Effective Teachers Using Performance on the Job, Washington, DC: The Brookings Institution.
- Hanushek, E. and S. Rivkin, Teacher Quality, in E. Hanushek and F. Welch, *Handbook of the Economics of Education*, Amsterdam: Elsevier.
- Hanushek, E., J. Kain, D. O'Brien and S. Rivkin, 2005, The market for teacher quality, NBER Working Paper 11154.
- Jacob, B.A. and L. Lefgren, 2004b, The impact of teacher training on student achievement: quasi-experimental evidence from school reform efforts in Chicago, *Journal of Human Resources* 39(1): 50–79.
- Jacob, B.A., 2002, Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools, NBER Working Paper 8968.
- Kane, T., J. Rockoff and O. Staiger, 2008, What does certification tell us about teacher effectiveness? Evidence from New York City, *Economics of Education Review*, 27(6): 615-631.
- Ladd, H.F., 1999, The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes, *Economics of Education Review*, 18, 1-16.
- Ladd, H.F. and R.P. Walsh, 2002, Implementing value-added measures of school effectiveness: Getting the incentives right, *Economics of Education Review*, 21, 1-17.
- Lavy, V., 2002a, Evaluating the effect of teachers' group performance incentives on pupil achievement, *Journal of Political Economy*, 110(6), 1286-1317.
- Lavy, V., 2002b, Paying for performance: The effect of teachers' financial incentives on students' scholastic outcomes, Mimeo, Universiteit van Jerusalem.
- Lazear, E.P. and S. Rosen, 1981, Rank-order tournaments as optimum labor contracts, *Journal of Political Economy*, 89(5), 841-864.

Lazear, E.P., 1999, Personnel economics: Past lessons and future directions, Presidential address to the Society of Labor Economists, San Francisco, May 1, 1998, *Journal of Labor Economics*, 17(2), 199-236.

Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink, 2007, The effect of extra funding for disadvantaged pupils on achievement, *Review of economics and statistics*, vol. 89, no. 4, pp. 721-736.

Loeb, S., M.E. Page, 2000, Examining the link between wages and student outcomes: The importance of alternative market opportunities and non-pecuniary variation, *Review of Economics and Statistics*, vol. 82, no. 3, pp. 393-408.

Muralidharan, K., V. Sundararaman, 2008, Teacher Performance Pay: Experimental Evidence from India, University of California, San Diego.

Rivkin, S., E. Hanushek and J. Kain, 2005, Teachers, schools and academic achievement, *Econometrica*, 73(2): 417-458.

Rockoff, J., B. Jacob, T. Kane and D. Staiger, 2008, Can you recognize an effective teacher when you recruit one?, NBER Working Paper 14485.

Rockoff, 2008, Does Mentoring Reduce Turnover and Improve Skills of New Employees? Evidence from Teachers in New York City, NBER Working Paper 13868.

Todd, P.E. and K.I. Wolpin, 2003, On the specification and estimation of the production function for cognitive achievement, *Economic Journal*, 113(485), F3-F33.

Wößmann, L., 2003a, Schooling Resources, Educational Institutions and Student Performance: the International Evidence, *Oxford Bulletin of Economics and Statistics*, 65, 117-170.

Speciaal onderwijs

Blok, H. and I. Breetvelt, 2002, *Adaptief onderwijs: betekenis en effectiviteit*, Amsterdam: SCO-Kohnstamm Instituut.

Cullin, J.B., 1999, The impact of fiscal incentives on student disability rates, NBER Working Paper 7173, June.

Cullen J.B. and S. G. Rivkin, 2001, The role of special education in school choice, in *The Economics of school choice*, NBER, Chapter 3 (NBER Reporter Summer 2001).

Hanushek E.A., J.F.Kain, S.G. Rivkin, 1998, Does special education raise academic achievement for students with disabilities? NBER Working Paper 6690, August.

Hanushek en Welch (ed), 2006, *Handbook of Economics of Education*, vol I en II OECD, 2004, *Equity in Education, Students with disabilities, learning difficulties and disadvantages*, Centre for Educational Research and Innovation, ISBN-92-64-10368-6.

Meijer, C.J.W. (Ed.), 2003, *Special Education across Europe in 2003. Trends in provision in 18 European countries*, Middelfart, Denmark: European Agency for Development in Special Needs Education.

Smeets, E.F.L. and J. Rispens, 2008, *Op zoek naar passend onderwijs. Overzichtsstudie van de samenhang naar regulier en speciaal (basis)onderwijs*, ITS/Radboud Universiteit Nijmegen.

Vermaas, J.C. and J. Van der Pluijm, 2005, *Brede evaluatie WSNS LGF en OAB*, Samenvattend overzicht van de onderzoeksrapporten en reviews, Den Haag: Ministerie van OCW.

Achterstandenbeleid

Angrist, J.D., E. Bettinger, E., Bloom, E., King and M. Kremer, 2002, Vouchers for private schooling in Columbia: Evidence from a randomized natural experiment, *American Economic Review*, 92(5):1535–1558.

Angrist, J.D. and V. Lavy, V., 2002, The effect of high school matriculation awards: Evidence from randomized trials, Working Paper Series 9389, NBER.

Camerer, C. F. and R.M. Hogarth, 1999, The effects of financial incentives in experiments: A review and capital-labor-production framework, *Journal of Risk and Uncertainty*, 19:7-42.

Dearden, L., C. Emmerson, C. Frayne, A. Goodman, H. Ichimura and C. Meghir, 2001, Education maintenance allowances: The first year - a quantitative evaluation, Research Report RR257, DfES.

Dearden, L., C. Emmerson, C., Frayne and C. Meghir, 2002, Education maintenance allowances: The first two years - a quantitative evaluation, Research Report RR352, DfES.

Dynarski, M., P. Gleason, A. Rangarajan, R. Wood, 1998, Impacts of Dropout Prevention Programs, A research report from the School Dropout Demonstration Assistance Program Evaluation.

Gneezy, U. and A. Rustichini, 2000, Pay enough or don't pay at all, *Quarterly Journal of Economics*, 110:791–810.

Hanushek, E.A., 1986, The economics of schooling: Production and efficiency in public schools, *Journal of Economic Literature*, 24:1141–1177.

Hanushek, E.A., 1996, School resources and student performance, In Burtless, G., editor, *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, The Brookings Institution.

Hoxby, C.M., 2000, The effects of class size on student achievement: New evidence from population variation, *Quarterly Journal of Economics*, 115(4).

Kremer, M., E. Miguel and R. Thornton, 2008, Incentives to Learn forthcoming in *Review of Economics and Statistics*.

Leuven, E., H. Oosterbeek and B. van der Klaauw, 2003, The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment, CEPR Discussion Paper 3921.

Ludwig, J. and D. Miller, 2007, Does Head Start improve children's life changes? Evidence from a regression discontinuity design, *Quarterly Journal of Economics*, vol 122, no. 1, pp 159-208.

Tymms, P., C. Merrell, J. Andor, K. Topping, A. Thurston and D. Miller, 2009, Improving attainment across a whole district: peer tutoring in a randomized controlled trial, Paper gepresenteerd op de AERA, San Diego, april 2009.

Excellentie

Angrist, J., D. Lang, and P. Oreopoulos, 2006, Lead Them to Water and Pay Them to Drink: An Experiment with Services and Incentives for College Achievement, NBER Working Paper 12790.

CPB, 2003, Studiefinanciering volgens het boekje, *Macro Economische Verkenning* 2004, Speciaal Onderwerp, Den Haag.

Dynarski, S., 2003, Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion, *American Economic Review*, 93(1), pp. 279-288.

Field, E., 2009, Educational debt burden and career choice: Evidence from a financial aid experiment at NYU Law School, *American Economic Journal Applied Economics*, 1(1), 1-21.

Hanushek, E.A., 2006, School Resources, in Hanushek, E.A. and Welch, F. (ed.), *Handbook of the Economics of Education*.

Hanushek, E.A. and L. Woessmann, 2007, The role of school improvement in economic development, NBER Working Paper 12832, NBER.

Hoxby, C., 2001, The return to attending a more selective college: 1960 to the present, Forum Futures, Exploring the futures of higher education, 2000 Papers.

Jacobs, B., 2002, An investigation of education finance reform, Den Haag: Centraal Planbureau.

Leuven, E., H. Oosterbeek, B. van der Klauw, 2009, The effect of financial rewards on students achievement: evidence from a randomized experiment, *Journal of the European Economic Association* (forthcoming).

Minne, B., M. Rensman, B. Vroomen, D. Webbink, 2007, Excellence for productivity, Den Haag: Centraal Planbureau.

School en omgeving

Sheldon, S., 2007, Improving student attendance with school, family, and community partnerships, *Journal of Educational Research*, 100(5), pp. 267-275.

Sheldon, S., J. Epstein, Involvement Counts: Family and community partnerships and mathematics achievement, *Journal of Educational Research*, 98(4), pp. 196-206.

Onderzoeksagenda

Cook, T. D., 2002, Randomized experiments in educational policy research. A critical examination of the reasons the educational evaluation community has offered for not doing them, *Educational Evaluation and Policy Analysis*, 24(3), 175-199.