

Op zoek naar bewijs

Op zoek naar bewijs

Evaluatieontwerpen onderwijsmaatregelen

Lex Herweijer en Monique Turkenburg

Sociaal en Cultureel Planbureau
Den Haag, september 2012

Het Sociaal en Cultureel Planbureau is ingesteld bij Koninklijk Besluit van 30 maart 1973.

Het Bureau heeft tot taak:

- a wetenschappelijke verkenningen te verrichten met het doel te komen tot een samenhangende beschrijving van de situatie van het sociaal en cultureel welzijn hier te lande en van de op dit gebied te verwachten ontwikkelingen;
- b bij te dragen tot een verantwoorde keuze van beleidsdoelen, benevens het aangeven van voor- en nadelen van de verschillende wegen om deze doeleinden te bereiken;
- c informatie te verwerven met betrekking tot de uitvoering van interdepartementaal beleid op het gebied van sociaal en cultureel welzijn, teneinde de evaluatie van deze uitvoering mogelijk te maken.

Het SCP verricht deze taken in het bijzonder bij problemen die het beleid van meer dan één departement raken.

De minister van Volksgezondheid, Welzijn en Sport is als coördinerend minister voor het sociaal en cultureel welzijn verantwoordelijk voor het door het SCP te voeren beleid. Over de hoofdzaken hiervan heeft hij/zij overleg met de minister van Algemene Zaken; van Veiligheid en Justitie; van Binnenlandse Zaken en Koninkrijksrelaties; van Onderwijs, Cultuur en Wetenschap; van Financiën; van Infrastructuur en Milieu; van Economische Zaken, Landbouw en Innovatie en van Sociale Zaken en Werkgelegenheid.

© Sociaal en Cultureel Planbureau, Den Haag 2012

SCP-publicatie 2012-22

Zet- en binnenwerk: Textcetera, Den Haag

Vertaling samenvatting: Julian Ross, Carlisle, Engeland

Omslagontwerp: Bureau Stijlzug, Utrecht

ISBN 978 90 377 0618 5

NUR 740

Voorzover het maken van reprografische verveelvoudigingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet 1912 dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 3060, 2130 KB Hoofddorp, www.repro-recht.nl). Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (art. 16 Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro).

Sociaal en Cultureel Planbureau

Parnassusplein 5

2511 VX Den Haag

Telefoon (070) 340 70 00

Fax (070) 340 70 44

Website: www.scp.nl

E-mail: info@scp.nl

De auteurs van SCP-publicaties zijn per e-mail te benaderen via de website. Daar kunt u zich ook kosteloos abonneren op elektronische attendering bij het verschijnen van nieuwe uitgaven.

Inhoud

Deel I	Onderzoek naar de effecten van het onderwijsbeleid	7
1	Aanleiding van deze studie	9
2	Evaluatieontwerpen	10
3	De effectevaluatie	11
4	De verklarende evaluatie	16
5	Kabinetsbeleid voor onderwijs	19
	Noot	20
6	Aanpak en reikwijdte van de ontwerpen	21
Deel II	Evaluatieontwerp voor vijftien onderwijsbeleidsmaatregelen	23
1	Prestatiebox primair onderwijs	25
2	Verplichte eindtoets en verplicht leerlingvolgsysteem in het primair onderwijs	33
	Noot	40
3	Prestatiebox voortgezet onderwijs	41
4	Aanscherping exameneisen in het voortgezet onderwijs	47
5	Prestatiebox middelbaar beroepsonderwijs	53
6	Introductie van entreeopleidingen in het mbo en afschaffing van de drempelloze instroom in mbo-2	57
7	Centrale examinering van taal en rekenen in het mbo	63
8	Standaardisering beroepsgerichte examens in het mbo	70
9	Versteving van de kennisbasis in het hbo middels landelijke toetsen	76
	Noot	81

10	Aanscherpen toezicht door de inspectie in het hoger onderwijs	82
11	Kwaliteit/prestatieafspraken en profiel in de bekostiging	87
12	Collegelddifferentiatie Noten	94 99
13	School aan zet	100
14	Invoering referentieniveaus in het primair, voortgezet en middelbaar beroepsonderwijs Noten	108 114
15	Leraren een masteropleiding Noten	115 123
Deel III Samenvatting en conclusie		125
1	Samenvatting van de evaluatieontwerpen	127
2	Kanttekeningen	129
3	Ten slotte	132
Literatuur		133
Publicaties van het Sociaal en Cultureel Planbureau		135

DEEL I

Onderzoek naar de effecten van het onderwijsbeleid

1 Aanleiding van deze studie

Het ministerie van o c w zet sinds een aantal jaren in op een versterking van de wetenschappelijke onderbouwing van het onderwijsbeleid. Nieuw beleid moet bij voorkeur zijn gebaseerd op wetenschappelijk gefundeerde inzichten over de effectiviteit van maatregelen (*evidence based*). Evaluatieonderzoek kan daaraan bijdragen, en bovendien aanknopingspunten bieden voor het bijstellen van reeds ingevoerd beleid. In het kader van het programma ‘Zicht op effectiviteit’ heeft de directie Kennis van het ministerie van o c w het s c p opdracht gegeven evaluatieontwerpen op te stellen voor een reeks van onderwijsmaatregelen, deels maatregelen uit het regeerakkoord van het kabinet-Rutte en deels van het vorige kabinet. Het betreft een tweede tranche van maatregelen; eerder maakten het Centraal Planbureau en de Erasmus Universiteit Rotterdam evaluatieontwerpen voor een eerste tranche van maatregelen (Van Elk et al. 2011). De evaluatieontwerpen kunnen in een latere fase dienen als uitgangspunt voor feitelijke evaluatie van de verschillende maatregelen.

Inzicht in de effectiviteit van maatregelen is voor het ministerie van o c w van belang in discussies met het onderwijsveld en met andere ministeries, waaronder dat van Financiën. Daarnaast is het belangrijk voor de doorrekening van onderwijsmaatregelen uit verkiezingsprogramma’s door het Centraal Planbureau (c p b). Bij die doorrekening geeft het c p b een schatting van de bijdrage aan de welvaart van verschillende maatregelen. Maatregelen waarvan het effect niet bekend is, kunnen hierin niet worden meegenomen. Bij de laatste doorrekening van verkiezingsprogramma’s was de trend zichtbaar dat politieke partijen in de richting werden gedrukt van interventies die goed zijn geëvalueerd. Dat heeft het risico in zich van een zekere verschraling van voorgesteld beleid. Als van een breder palet van maatregelen bekend is wat de effecten zijn, is dit risico minder groot.

2 Evaluatieontwerpen

Dit rapport bevat evaluatieontwerpen voor vijftien onderwijsmaatregelen. Deze maatregelen hebben betrekking op het primair onderwijs, het voortgezet onderwijs, het middelbaar beroepsonderwijs en het hoger onderwijs. In de evaluatieontwerpen geven we een overzicht van de achtergrond, de inhoud, de doelstelling en de veronderstellingen die aan de maatregel ten grondslag liggen. Daarna schetsen we hoe de evaluatie van de maatregel kan worden aangepakt, waarbij we onderscheid maken tussen een effectevaluatie en een verklarende evaluatie.

In een *effectevaluatie* staat de vraag centraal of een maatregel het beoogde effect sorteert, een *verklarende evaluatie* richt zich op het achterhalen van de mechanismen waaruit het optreden van effecten kan worden verklaard. De gedachte is dat het de bruikbaarheid en geloofwaardigheid van evaluatieresultaten ten goede komt als een evaluatie niet alleen laat zien dat een maatregel effect heeft, maar ook toont hoe het effect tot stand komt. Van sommige maatregelen zal het effect lastig met zekerheid zijn vast te stellen, omdat hiervoor geen hard experimenteel bewijs kan worden verkregen. De uitkomsten van de verklarende evaluatie kunnen in die gevallen waardevol zijn, omdat ze het minder zekere beeld dat uit de effectevaluatie naar voren komt kunnen ondersteunen.

Maatregelen kunnen naast bedoelde effecten ook neveneffecten teweegbrengen. Het is belangrijk om hiervoor bij het ontwerpen van de evaluaties oog te hebben, zeker als deze neveneffecten negatief zijn. Een neveneffect kan ook zijn dat verschillende nieuwe maatregelen op elkaar ingrijpen. Door interferentie met andere maatregelen kan het effect van een nieuwe maatregel anders uitpakken dan bij afzonderlijke invoering.

Evaluatie kan deel uitmaken van de ontwikkeling en invoering van nieuw beleid. Daarbij moet de aanpak van de evaluatie bij voorkeur worden afgestemd op het ontwikkelingsstadium waarin een interventie zich bevindt (Van Yperen en Bijl 2009). Op een nieuw terrein is eerst een periode van verkennend onderzoek, ontwikkelwerk en praktijkervaring nodig om een beeld te krijgen van wat werkt, en waarom en hoe. Als die fase achter de rug is, kunnen desgewenst experimenten met controlegroepen worden gestart om 'hard' bewijs te krijgen van de effecten van een maatregel (Onderwijsraad 2006; HM Treasury 2011). Het starten van experimenten met controlegroepen is niet zinvol als er nog geen theorie is die specificeert wat werkzame factoren zijn en als er nog geen voorlopig onderzoeksmateriaal is dat indicaties geeft van mogelijke effecten. Afhankelijk van de stand van de kennisontwikkeling op een bepaald gebied zijn dus verschillende gradaties van bewijsvoering rondom nieuwe maatregelen mogelijk. In de praktijk zal een gefaseerde aanpak – waarbij eerst op beperkte schaal ervaring wordt opgedaan met een maatregel, vervolgens experimenten worden uitgevoerd om de effecten ervan aan te tonen en de maatregel uiteindelijk sectorbreed wordt ingevoerd – lang niet altijd mogelijk zijn. Het is niet uitzonderlijk dat maatregelen worden ingevoerd in een hele sector zonder voorafgaande pilots of experimenten. Dit heeft wel gevolgen voor de opzet van evaluatieonderzoek.

3 De effectevaluatie

In de evaluatieontwerpen maken we onderscheid tussen de effectevaluatie en de verklarende evaluatie. Een effectevaluatie richt zich op de vraag of de effecten die van een maatregel worden verwacht zich in de praktijk inderdaad voordoen. Om bewijs te krijgen dat een maatregel het beoogde effect heeft, moeten de resultaten van de maatregel bij een ‘experimentele’ groep, die het beleid ondergaat, worden vergeleken met de resultaten bij een controlegroep, die niet aan de maatregel wordt blootgesteld. Het centrale probleem hierbij is de geloofwaardigheid van de controlegroep: geeft de controlegroep een goed beeld van hoe de experimentele groep ervoor zou hebben gestaan als de interventie niet zou hebben plaatsgevonden (Webbink 2011)? Om die vraag goed te kunnen beantwoorden is het belangrijk dat de controlegroep en de experimentele groep vergelijkbaar zijn. Niet-willekeurige (zelf)selectie bij verdeling van actoren (scholen, leerkrachten, leerlingen) over de experimentele en de controlegroep kan de vergelijkbaarheid van de experimentele en controlegroep aantasten. Als een kenmerk dat van invloed is op de verdeling van actoren over de experimentele en de controlegroep ook invloed heeft op het resultaat dat met een interventie wordt beoogd, bestaat het risico dat ten onrechte een effect aan de interventie wordt toegeschreven. Denk hierbij bijvoorbeeld aan een experiment met een nieuwe aanpak waarvoor met name deskundige en gemotiveerde leerkrachten zich aanmelden. Vanwege het verschil in deskundigheid en motivatie zijn de resultaten die deze leerkrachten realiseren met de nieuwe aanpak niet goed te vergelijken met de resultaten die een controlegroep van doorsnee-leerkrachten behaalt met de oude aanpak. Het risico is dat geconcludeerd wordt dat de nieuwe aanpak beter werkt dan de oude aanpak, terwijl het verschil in resultaten in werkelijkheid het gevolg is van de grotere deskundigheid en de betere motivatie van de leerkrachten in de experimentele groep.

Random gecontroleerde experiment

Er zijn verschillende mogelijkheden om het probleem van vertekening door selectie-effecten bij de vergelijking tussen een experimentele en een controlegroep tegen te gaan. De ‘gouden standaard’ is het *random gecontroleerde experiment*, waarbij actoren (scholen, leerkrachten, leerlingen) volgens het toeval aan een experimentele en een controlegroep worden toegewezen (Bosker 2009). Random gecontroleerde experimenten zijn om uiteenlopende redenen tot op heden een zeldzaamheid in het Nederlandse onderwijsbeleid. Het onthouden van een behandeling aan een controlegroep wordt soms namelijk als onethisch bestempeld, en experimenten vergen tijd, geld en bereidwilligheid van het onderwijsveld.

Quasi-experimentele opzet

Als een random gecontroleerd experiment geen optie is, moeten de mogelijkheden van een quasi-experimentele opzet worden verkend. Ook dit is een opzet die uitgaat van vergelijking tussen een experimentele en een controlegroep, maar anders dan bij het

random gecontroleerde experiment wordt de samenstelling van beide groepen hier niet gestuurd door de onderzoekers.

Er kan onderscheid worden gemaakt tussen een quasi-experimentele opzet met een gelijkwaardige experimentele en controlegroep en een quasi-experimentele opzet met een niet gelijkwaardige experimentele en controlegroep (HM Treasury 2011).

Quasi-experimentele opzet met gelijkwaardige experimentele en controlegroep
Soms kan gebruik worden gemaakt van variaties in de praktijk, waardoor bij toeval goed vergelijkbare experimentele en controlegroepen ontstaan. Een voorbeeld zijn lotingen die bepalen of actoren wel of niet aan aanmerking komen voor een maatregel. De groep die is ingeloot, fungeert dan als experimentele groep; de groep die is uitgeloot, vormt de controlegroep.

Een andere mogelijkheid is dat wordt gebruikgemaakt van een gefaseerde invoering van een maatregel. De groep die de interventie in een latere fase ondergaat, fungeert daarbij als controlegroep voor de groep die als eerste aan de beurt is. Daarbij is het wel belangrijk dat de fasering min of meer toevallig tot stand komt, en niet is gebaseerd op bijvoorbeeld de ernst van de problemen waarop de interventie zich richt of op het enthousiasme van de betrokkenen.

Een nog weer ander mogelijkheid is dat wordt gebruikgemaakt van beslissingsregels die bij een indicatiestelling worden gehanteerd, waardoor sommige actoren net wel en andere net niet in aanmerking komen voor een maatregel of voorziening. De actoren die zich direct aan weerszijden van deze scheidslijn bevinden, lijken sterk op elkaar en kunnen dus fungeren als experimentele groep (komt net wel in aanmerking voor de maatregel) en controlegroep (komt net niet in aanmerking voor de maatregel). Een beperking is dat de vergelijking tussen experimentele en controlegroep hierbij alleen betrekking heeft op actoren dicht bij deze scheidslijn. De resultaten van de vergelijking kunnen niet zonder meer worden vertaald naar actoren die zich verder van de scheidslijn bevinden (HM Treasury 2011).

Quasi-experimentele opzet met niet gelijkwaardige experimentele en controlegroep

Als het niet aannemelijk is dat de experimentele en controlegroep die in de praktijk zijn ontstaan gelijkwaardig zijn, moet bij de vergelijking tussen beide groepen rekening worden gehouden met mogelijke vertekening door selectie-effecten. Dat kan op verschillende manieren.

Een eerste optie is de zogenoemde *difference-in-difference* benadering. Deze gaat uit van een vergelijking van trends bij de experimentele en de controlegroep, waarbij een verschil in trend tussen beide groepen een aanwijzing is voor het effect van de maatregel. Onderkend wordt dat het *niveau* van een uitkomstmaat in een niet-gelijkwaardige experimentele en controlegroep niet vergelijkbaar is. Factoren die samenhangen met de verdeling van actoren over de experimentele en de controlegroep kunnen namelijk ook van invloed zijn op het niveau van de uitkomstmaat. Denk bijvoorbeeld aan het eindniveau dat studenten halen in een uitdagend studieprogramma waarvoor alleen de meest

getalenteerde studenten worden geselecteerd. Als studenten in zo'n programma een hoog eindniveau halen, zal dat mede een gevolg zijn van deze selectie. Verondersteld wordt echter dat de *trend* in de uitkomstmaat in de experimentele en de controlegroep wel vergelijkbaar zou zijn als de experimentele groep niet aan de maatregel zou zijn blootgesteld. Een waargenomen verschil in trend tussen de experimentele en de controlegroep na invoering van de maatregel wijst dan op een effect van de maatregel. De aanname van een gelijke trend in de experimentele en de controlegroep (de zogenoemde *common-trends*assumptie) is essentieel. De geldigheid van deze aanname kan niet worden aangetoond, maar hooguit plausibel worden gemaakt. Dat is bijvoorbeeld mogelijk door de trend in beide groepen in de periode voorafgaand aan de interventie te vergelijken (HM Treasury 2011).

Een alternatieve aanpak is dat bij de vergelijking tussen de experimentele en de controlegroep wordt gecorrigeerd voor relevante verschillen tussen beide groepen. Dat kan door middel van *regressiemodellen*, waarin zo veel mogelijk kenmerken zijn opgenomen die de vergelijking tussen de beide groepen kunnen vertekenen. Het kan ook door *matching*. Daarbij worden aan de hand van relevante kenmerken subgroepen uit de experimentele en de controlegroep geselecteerd die wel goed vergelijkbaar zijn (HM Treasury 2011).

Zowel bij statistische correctie als bij matching is het nooit zeker dat voldoende rekening is gehouden met relevante factoren. Omdat niet valt uit te sluiten dat ook niet-geobserveerde kenmerken van invloed zijn op gevonden verschillen tussen de experimentele en de controlegroep is het bewijs bij deze aanpak minder 'hard'.

In de praktijk kunnen verschillende methoden worden gecombineerd. Een trendvergelijking kan worden toegespitst op subgroepen uit de experimentele en de controlegroep die vergelijkbaar zijn op relevante kenmerken. En omgekeerd wint een vergelijking tussen gematchte groepen aan kracht als de vergelijking zich richt op trends.

Integraal ingevoerd beleid; verschillen in beleidsintensiteit

Bij maatregelen die integraal worden ingevoerd, is een vergelijking tussen een experimentele en een – al dan niet equivalente – controlegroep in principe niet mogelijk. In de praktijk kan de intensiteit van integraal doorgevoerde maatregelen echter variëren. Als dat het geval is, kan de effectevaluatie aansluiten bij deze variatie in intensiteit. Actoren (scholen, leerkrachten, leerlingen) die weinig worden geraakt door een maatregel fungeren hierbij als 'controlegroep' en actoren die sterk worden geraakt vormen de 'experimentele' groep.

De uitgangssituatie van beide groepen zal niet vergelijkbaar zijn, zoals bij een random gecontroleerd experiment of een quasi-experimentele opzet met een vergelijkbare experimentele en controlegroep. Op basis van een trendvergelijking (*difference-in-difference*benadering) kan niettemin worden afgeleid in hoeverre het beleid effect heeft. De aanname van gelijke trend in beide groepen als het beleid niet zou zijn ingevoerd, is hierbij weer essentieel.

Ook de controlegroep staat in enige mate bloot aan het beleid. Een verschil in trend tussen de experimentele en de controlegroep wijst daarom niet op het effect van het beleid als zodanig, maar op het effect van verschil in intensiteit van beleid. Zijn de veranderingen in uitkomsten groter bij de groep die sterk door het beleid wordt geraakt dan bij de groep die er weinig door wordt geraakt?

Integraal ingevoerd beleid: populatievergelijking

Als bij een integraal ingevoerde maatregel geen groepen zijn te vinden die van elkaar verschillen in de mate waarin ze aan het beleid zijn blootgesteld, is er de mogelijkheid van een populatiebrede vergelijking voor en na de invoering van de maatregel. Op basis van een trend uit het verleden die wordt doorgetrokken naar de toekomst kan een voorspelling worden gedaan, die vervolgens fungeert als *counterfactual* voor de waargenomen toestand na de invoering van de maatregel. Een afwijking ten opzichte van de voorspelde trend zou dan wijzen op een effect van de maatregel. Denk hierbij bijvoorbeeld aan een vergelijking van de loopbaan van complete cohorten leerlingen of studenten voor en na de invoering van een maatregel.

Een probleem van een dergelijke trendvergelijking zonder controlegroep is dat allerlei omstandigheden die van invloed zijn op de uitkomst die met een maatregel wordt beoogd eveneens van jaar op jaar kunnen veranderen. Veranderingen in uitkomsten kunnen daardoor niet zonder meer worden toegeschreven aan de maatregel. Als bijvoorbeeld de invoering van een maatregel om het voortijdig schoolverlaten terug te dringen samenvalt met een relevante verandering in de economische conjunctuur, is een daling van het aantal voortijdig schoolverlaters geen bewijs van het effect van de maatregel. Hooguit kan het effect daarvan plausibel worden gemaakt door de invloed van veranderende omstandigheden zo veel mogelijk uit te filteren.

Als gelijktijdig verschillende maatregelen integraal worden ingevoerd, is evaluatie van het effect van de afzonderlijke maatregelen niet mogelijk. Ook als effectevaluatie van integraal ingevoerd maatregelen niet mogelijk is, blijft het belangrijk om de mate van doelrealisatie te 'monitoren'.

Gradaties van bewijsvoering

Het bewijs van het effect van maatregelen is niet bij alle opzetten even 'hard'. Het random gecontroleerde experiment en de quasi experimentele opzetten met gelijkwaardige experimentele en controlegroepen bieden de beste garanties voor een robuuste effectmeting. Dankzij de willekeurige verdeling van actoren over de experimentele en de controlegroep worden de met deze opzetten gevonden effecten van maatregelen niet vertekend door selectie-effecten.

Bij quasi-experimentele opzetten met een niet gelijkwaardige experimentele en controlegroep is het bewijs minder hard. Bij een trendvergelijking van niet gelijkwaardige experimentele en controlegroepen komt het aan op de geloofwaardigheid van de aanname van gelijke trend. Deze aanname kan, zoals gezegd, niet worden bewezen, maar bijvoorbeeld wel worden onderbouwd op basis van trends in het verleden. Bij statistische correctie voor verschillen tussen een niet gelijkwaardige experimentele en

controlegroep en bij matching blijft het onzeker of afdoende rekening is gehouden met relevante verschillen tussen de experimentele en de controlegroep.

Als, ten slotte, bij gebrek aan een controlegroep alleen vergelijking van de totale populatie op opeenvolgende momenten mogelijk is, zijn in de regel geen harde conclusies over effecten van maatregelen te trekken.

In de praktijk kan het soms beter zijn om te kiezen voor een minder harde variant van effectevaluatie. Zo vereist een random experiment een experimentele en een controlegroep van voldoende omvang. Als slechts een beperkt aantal actoren aan het experiment kan meedoen, bijvoorbeeld vanwege budgettaire redenen, is er een gerede kans dat een willekeurige toedeling van actoren over de experimentele en de controlegroep niet leidt tot vergelijkbare groepen. Het kan dan beter zijn om een opzet te kiezen waarbij de experimentele en de controlegroep vooraf worden gematcht op relevante kenmerken, ook al is het nooit helemaal zeker dat daarbij met alle relevante kenmerken rekening wordt gehouden.

4 De verklarende evaluatie

In de vorige paragraaf is stilgestaan bij de mogelijkheden om het effect van een beleidsmaatregel vast te stellen en zijn de verschillende evaluatiemogelijkheden in een min of meer hiërarchische volgorde gepresenteerd. Bij dit type evaluaties betreft het een oplossing voor het attributieprobleem: het kunnen toeschrijven van effecten aan een interventie/maatregel. Afhankelijk van de gekozen opzet kan een causale relatie worden gelegd tussen de interventie en de uitkomsten.

Dergelijk evaluatieonderzoek maakt op zichzelf echter nog niet duidelijk waarom de effecten wel of niet of niet volledig optraden of waarom er ongewenste neveneffecten zijn. Om daarop zicht te krijgen is aanvullende informatie nodig. In deze paragraaf staan we stil bij een evaluatie die naast een effectmeting plaatsvindt: de verklarende evaluatie.

De verklarende evaluatie richt zich op de implementatie van de beleidsmaatregel. Vragen die hierbij centraal staan zijn: Hoe en in welke context zijn de maatregelen uitgevoerd? Welke aanpak heeft men gekozen en waarom? Wat werkte goed volgens de betrokkenen en waarom was dat zo? Wat werkte minder goed en hoe kwam dat? Wat zijn de ervaringen van de doelgroep met de maatregelen, wat was waardevol en wat veroorzaakte juist problemen? Was er daarbij verschil tussen de diverse groepen? Welke onbedoelde neveneffecten zijn opgetreden, en werd de maatregel geïntervenieerd door ander beleid? Een verklarende evaluatie levert dus bijzonder nuttige informatie op over de manier waarop de maatregel in de praktijk functioneert en over de manier waarop de kwaliteit van het beleid kan worden verbeterd. Daarnaast biedt zij informatie voor het ontwerpen voor toekomstig beleid (HM Treasury 2011). De verklarende evaluatie is overwegend inventariserend en beschrijvend van aard.

De verklarende evaluatie wordt overigens meestal aangeduid met de term procesevaluatie. Wij kiezen in deze rapportage voor de term *verklarende evaluatie* om daarmee te benadrukken dat de procesevaluatie is gericht op het vinden van *de mechanismen* die kunnen verklaren waarom een beleidsmaatregel wel of niet het beoogde effect teweegbrengt. Van sommige maatregelen valt het effect echter niet goed vast te stellen, bijvoorbeeld omdat er geen vergelijkingsgroep is te vinden. Een verklarende evaluatie in de letterlijke zin van het woord is in zo'n geval niet mogelijk. Het verklaren waarom effecten optreden, veronderstelt dat bekend is of een maatregel effect heeft. Wel kan dan worden nagegaan of de invoering van de maatregel in overeenstemming is met theoretische inzichten en praktijkkennis over de werkzaamheid van de desbetreffende interventie.

Het vertrekpunt voor de verklarende evaluatie vormt *de beleidstheorie*: het geheel aan veronderstellingen over de relatie tussen de ingezette middelen van het beleid, de doelen van het beleid en het gedrag van degenen op wie het beleid gericht is. Hoe wordt het beleid geacht te werken, wat waren de assumpties van de beleidsmakers? Na reconstructie van de beleidstheorie wordt deze getoetst, wat tijdens elke beleidsfase mogelijk is.

Verklarende evaluatie kan dus worden gezien als een vorm van hypothesetoetsend onderzoek en kan *ex ante*, *ex durante* of *ex post* worden toegepast. In de beleidsevaluatie moet worden afgewogen op welk moment de verklarende evaluatie moet starten. Is dat bijvoorbeeld kort na het van kracht worden van de maatregel, zodat eerste bevindingen over de implementatie eventueel nog tot een bijstelling van het beleid kunnen leiden? Of juist in een later stadium, zodat ook de doorwerking van de maatregel kan worden onderzocht?

Het blijkt in de praktijk nog niet eenvoudig om de werkzame mechanismen van een interventie te achterhalen. De verschillende maatregelen zijn immers niet altijd even helder afgeperkt of eenduidig. Bovendien worden de maatregelen uitgevoerd in een *complexe, gelaagde structuur* met verschillende partijen (bv. sectorraden, schoolbesturen, directies van scholen, leerkrachten en gemeenten). Hierdoor kunnen verschillen ontstaan tussen het politieke besluit en de door uitvoerende instanties geleverde beleidsprestaties, de zogenoemde beleidsafwijkingen (Torenvlied 1996). Een belangrijke voorwaarde voor een succesvolle invoering van nieuwe maatregelen is dat het handelen van de partijen die op de verschillende niveaus bij de invoering zijn betrokken onderling voldoende is afgestemd, de zogenoemde *policy alignment of goal alignment* (Hofman et al. 2008; Waslander 2007). Een mogelijke verklaring voor het uitblijven van het beoogde effect kan inconsistentie zijn in doelen op de verschillende uitvoerende niveaus of gebrek aan overeenstemming tussen actoren. Het uitvoeren van een beleidsmaatregel omdat een externe partij dat verlangt, leidt niet zonder meer tot een succesvolle implementatie. Interne motivatie en een zekere mate van 'eigenaarschap' bij de betrokken actoren zijn cruciale voorwaarden om zich voor het beleidsdoel in te zetten.

In de verklarende analyse is het daarom van belang de relevante 'lagen' (en daarbinnen actoren) te onderscheiden, en per 'laag' de juiste vragen te formuleren om de mechanismen te achterhalen die het effect teweegbrengen of doen uitblijven. Zijn de invoering en de reacties van betrokkenen in overeenstemming met de beleidstheorie? Wat zijn de motieven van de verschillende actoren om zich wel of niet (volledig) voor het beleidsdoel in te zetten?

De verklarende evaluatie is dus gericht op het begrijpen hoe de implementatie en de uitvoering van het beleid plaatsvinden en om zicht te krijgen op factoren en mechanismen die de effectiviteit van het beleid bevorderen of hinderen. Het is daarbij belangrijk om zicht te krijgen op de verschillende *contexten* waarin de maatregel wordt in- en uitgevoerd. Het werkzame of causale mechanisme van een beleidsmaatregel kan immers afhangen van de context waarin deze plaatsvindt, en dus per context verschillen. De context is soms ook bepalend voor de specifieke invulling van een breed geformuleerde beleidsdoelstelling. Binnen zo'n bredere doelstelling krijgen instellingen vaak de mogelijkheid om alleen die maatregelen te kiezen die passen bij hun specifieke situatie. Daarmee neemt de contextgebondenheid van een maatregel toe en laat het beleid zich niet langer als een generieke maatregel evalueren.

De verklarende evaluatie heeft niet een vooraf vaststaand format. De invulling van de verklarende evaluatie hangt af van de te onderzoeken maatregel, de actoren die daarbij betrokken zijn, het tijdspad dat ermee gemoeid is, de mate van contextgebondenheid en bovenal: de vragen waarop de verklarende evaluatie een antwoord moet bieden. Ook de keuze voor verschillende onderzoeksactiviteiten en methoden hangt af van de vragen waarop de verklarende evaluatie antwoord moet geven.

In zijn algemeenheid kunnen we stellen dat *triangulatie* als onderzoeksstrategie de betrouwbaarheid van waarnemingen vergroot. Triangulatie wordt vaak toegepast in kwalitatief onderzoek en wil zeggen dat verschillende ‘metingen’ worden verricht vanuit verschillende invalshoeken. Triangulatie kan betrekking hebben op het systematisch vergelijken van verschillende cases, op het gebruik van verschillende databronnen of -verzamelingen, op verschillende theorieën ter verklaring of op verschillende analyse-technieken. Triangulatie kan dus verschillende vormen aannemen, die elkaar overigens niet uitsluiten. Theoretische triangulatie kan bijvoorbeeld worden gecombineerd met methodische triangulatie, datatriangulatie of onderzoekstriangulatie. Door een bepaald verschijnsel vanuit meerdere hoeken te belichten neemt de geldigheid van uitspraken over het onderzochte toe.

Triangulatie heeft hier als doel een zo compleet mogelijk, diepgaand inzicht te krijgen van het beleid in de praktijk.

Als in het verklarende onderzoek gekozen wordt voor een aanpak met casestudies, is het bijvoorbeeld zaak dat deze ‘variatiedekkend’ zijn. De resultaten van een *survey* onder betrokkenen naar hun ervaring met het nieuwe beleid wint bijvoorbeeld aan kracht door in *face to face* interviews om aanvullende informatie te vragen. Een beleidsmaatregel die sterk van invloed is op het groepsproces op een school zou in een verklarende evaluatie kunnen worden onderzocht door middel van participerende observaties in heel uiteenlopende cases, eventueel aangevuld met informatie uit werkplannen en verslagen.

We hopen hiermee duidelijk te hebben gemaakt dat er niet één (juiste) manier is om een verklarende evaluatie uit te voeren. In deel II zullen we per maatregel aangeven wat de relevante onderzoeksvragen voor de verklarende evaluatie zijn en hoe deze in het verklarend onderzoek (het beste) kunnen worden beantwoord. Ook bij maatregelen waarbij effectevaluatie niet mogelijk is, doen we een voorstel voor een verklarende evaluatie. In die gevallen kan in de verklarende evaluatie worden nagegaan of de invoering in overeenstemming is met theoretische inzichten en praktijkkennis over de desbetreffende interventie.

5 Kabinetsbeleid voor onderwijs

Het afgelopen jaar heeft het kabinet de uitgewerkte beleidsdoelen uit het regeerakkoord gepresenteerd. Het betreft actieplannen waarin per sector de beleidsdoelen zijn weergegeven (ocw 2011a t/m e). De grote lijn van het beleid, over de afzonderlijke sectoren heen, heeft betrekking op het garanderen van onderwijskwaliteit en op het verbeteren van de leerprestaties – ook aan ‘de bovenkant’, bij de excellente of hoogbegaafde leerlingen en studenten. De maatregelen hebben betrekking op het borgen van de kwaliteit van het onderwijs door bijvoorbeeld meer eisen te stellen aan het opleidingsniveau en de werkwijze van docenten (opbrengstgericht), aan de kwaliteit van examinering en aan de resultaten die leerlingen behalen. Daarnaast zijn er maatregelen die moeten zorgen voor profilering van instellingen en meer variëteit in opleidingen.

In dit rapport schetsen we een evaluatieontwerp voor een reeks van maatregelen uit de verschillende actieplannen en uit al lopend beleid. Het beleid omvat uiteraard veel meer maatregelen, maar deze opdracht beperkt zich tot de vijftien maatregelen die hieronder worden omschreven. Sommige maatregelen of interventies zijn beperkt tot een specifieke sector – bijvoorbeeld het voortgezet onderwijs – andere maatregelen zijn sectordoorsnijdend – zoals de invoering van referentieniveaus op het gebied van taal en rekenen.

Primair onderwijs

- 1 Prestatiebox¹ primair onderwijs;
- 2 Verplichte eindtoets en verplicht leerlingvolgsysteem in het primair onderwijs;

Voortgezet onderwijs

- 3 Prestatiebox voortgezet onderwijs;
- 4 Aanscherping exameneisen in het voortgezet onderwijs;

Middelbaar beroepsonderwijs

- 5 Prestatiebox middelbaar beroepsonderwijs;
- 6 Introductie van entreeopleidingen in het mbo en afschaffing van de drempelloze instroom in mbo-2;
- 7 Centrale examinering van taal en rekenen in het mbo;
- 8 Standaardisering beroepsgerichte examens in het mbo;

Hoger onderwijs

- 9 Versteving van de kennisbasis in het hbo middels landelijke toetsen;
- 10 Aanscherpen toezicht door de inspectie in het hoger onderwijs;
- 11 Kwaliteit/prestatieafspraken en profiel in de bekostiging;
- 12 Collegegelddifferentiatie;

Primair en voortgezet onderwijs

- 13 'School aan zet' (een programma om scholen en schoolbesturen in het primair en voortgezet onderwijs te ondersteunen bij het realiseren van de doelstellingen uit de prestatiebox);

Primair, voortgezet onderwijs en middelbaar beroepsonderwijs

- 14 Invoering van de referentieniveaus voor taal en rekenen;
- 15 Leraren een masteropleiding.

Noot

- 1 Onder 'prestatiebox' wordt verstaan: extra middelen voor scholen/ instellingen om te werken aan een reeks van beleidsdoelstellingen, met ruimte voor schoolbesturen om zelf accenten te leggen.

6 Aanpak en reikwijdte van de ontwerpen

Als voorbereiding op de evaluatieontwerpen hebben wij tussen 12 december 2011 en 13 januari 2012 gesprekken gevoerd met medewerkers van de betrokken beleidsdirecties (primair onderwijs, voortgezet onderwijs, leraren, beroeps- en volwasseneneducatie en hoger onderwijs). Daarbij waren tevens medewerkers van de directie Kennis aanwezig. Het doel van deze gesprekken was de beleidsmaatregel, de context waarin de maatregel plaatsvindt en de laatste stand van zaken in de beleidsontwikkeling zo goed mogelijk in beeld te brengen. Ook is bepaald in welk stadium de invoering van het beleid zich bevindt. Is in pilots al ervaring opgedaan met een maatregel en is er nu sprake van een bredere of integrale invoering van een maatregel? Is er eventueel nog ruimte voor het starten van pilots of experimenten? De verschillende beleidsdirecties hebben in de gesprekken aangegeven wat een beleidsevaluatie aan zinvolle informatie voor de directie zou moeten opleveren. Verder hebben onderzoekers en beleidsmedewerkers getracht om globaal af te tasten wat de evaluatiemogelijkheden van een specifieke maatregel zijn. Naast de gesprekken zijn relevante beleidsdocumenten bestudeerd.

Reikwijdte van de ontwerpen

Dit rapport betreft evaluatieontwerpen op hoofdlijnen: de ontwerpen geven geen volledig uitgewerkte blauwdruk van evaluatieonderzoeken voor de verschillende maatregelen. Als maatregelen daadwerkelijk worden geëvalueerd, moeten de ontwerpen verder worden uitgewerkt tot een volledige onderzoeksopzet. Het streven is om waar mogelijk gebruik te maken van reeds beschikbare gegevens, zodat de belasting van scholen, docenten, leerlingen en studenten zo gering mogelijk blijft. In de ontwerpen geven we aan welke bestaande gegevens benut kunnen worden. Bij het verder uitwerken van ontwerpen zal nader onderzoek naar de beschikbaarheid en de bruikbaarheid van gegevens noodzakelijk zijn.

Elk ontwerp geeft een beschrijving van de veronderstellingen achter een maatregel, van de beoogde effecten en van eventuele neveneffecten. We beperken ons daarbij tot de hoofdlijnen: een per maatregel volledig uitgewerkte reconstructie van de beleidstheorie was niet mogelijk binnen het bestek van deze opdracht. De verkenning van mogelijke neveneffecten is gebaseerd op kennis van het terrein en informatie uit de gesprekken die zijn gevoerd met de medewerkers van de beleidsdirecties.

Een uitgebreide inventarisatie van literatuur over de effecten en mogelijke neveneffecten van maatregelen was niet mogelijk gegeven de tijd die beschikbaar was voor het opstellen van de ontwerpen. Na de beschrijving van de maatregel volgt de opzet van de effectevaluatie en de verklarende evaluatie.

Bij het opstellen van de evaluatieontwerpen zijn we in de regel uitgegaan van de wijze waarop het beleid wordt ingevoerd. Als niet is voorzien in pilots of experimenten hebben we dat beschouwd als een gegeven. Bij enkele ontwerpen is daar in overleg met de opdrachtgever van afgeweken en doen we een voorstel voor een experiment.

DEEL II

Evaluatieontwerp voor vijftien onderwijsbeleidsmaatregelen

1 Prestatiebox primair onderwijs

1.1 Achtergrond en aanleiding

In de actieplannen van het kabinet voor de verschillende sectoren van het onderwijs zijn maatregelen aangekondigd om het prestatieniveau in de kernvakken te verhogen en excellentie te bevorderen. Geconstateerd wordt dat het Nederlandse onderwijs internationaal gezien vrij goed presteert, maar dat er indicaties zijn dat het beter kan. De zwakke leerling in het Nederlandse onderwijs doet het weliswaar relatief goed, maar de prestaties van de betere leerlingen blijven achter bij die van de betere leerlingen in andere landen. Daarnaast laat internationaal onderzoek zien dat de prestaties van leerlingen in het Nederlandse basis- en voortgezet onderwijs stagneren of zelfs enigszins achteruitgaan, terwijl die van leerlingen in een aantal andere landen zijn verbeterd. Scholen moeten daarom de komende jaren meer opbrengstgericht gaan werken en meer aandacht gaan besteden aan de beste leerlingen, zodat prestaties over de hele linie verbeteren en er meer uit de beste leerlingen wordt gehaald.

Nu nog worden scholen en instellingen behalve via de lumpsum ook gefinancierd via allerlei afzonderlijke subsidies. Deze vorm van financiering is vaak onzeker en bureaucratisch. Het kabinet wil de bekostiging van het onderwijs vereenvoudigen en de regels voor scholen verminderen. Een uitwerking daarvan is het reduceren van het aantal subsidies en het invoeren van een vorm van beloning voor te leveren prestaties: de prestatiebox. Deze vorm van financiering zal behalve in het primair onderwijs (po) ook worden ingevoerd in het voortgezet onderwijs (vo) en het middelbaar beroepsonderwijs (mbo). Landelijke beleidsdoelen en beleidsambities zijn doorgaans niet één op één naar schoolniveau overdraagbaar – daarvoor dient een vertaalslag te worden gemaakt. Scholen verschillen bovendien soms sterk in hun uitgangssituatie. Voor de ene school zal de urgentie zich op een van de landelijke beleidsdoelen te richten groter zijn dan voor een andere school. Een generieke toedeling van subsidie voor een bepaald doel is daardoor niet altijd even adequaat. De prestatiebox is een min of meer nieuwe sturingswijze die enerzijds globaal is (geen oormerking, toevoeging aan de lumpsum), maar anderzijds wel gekoppeld wordt aan de specifieke beleidsdoelen van het kabinet. Landelijke sturing vindt plaats door de prestatiebox te koppelen aan doelen waaraan in eerste instantie de sector, maar uiteindelijk toch de individuele scholen en instellingen zich verbinden.

1.2 Inhoud

De prestatiebox is een vorm van financiering bedoeld om de prestaties van leerlingen, de kwaliteit van het onderwijs en van de leerkrachten en schoolleiders te verbeteren. Schoolbesturen in het po ontvangen via de prestatiebox extra middelen om de beleidsambities te realiseren. De prestatiebox is feitelijk een ‘lege’ regeling. Op sectorniveau zijn – in dit geval met PO-Raad – meerjarige afspraken gemaakt over de doelen waarvoor de prestatiebox concreet zal worden ingezet gedurende de looptijd van het Actieplan, de

periode 2012-2015. Het bestuursakkoord tussen de minister en de PO-Raad is op 17 januari 2012 ondertekend.

De ambities uit het bestuursakkoord voor de periode 2012-2015 voor het po zijn de volgende:

- 1 Scholen werken opbrengstgericht.
- 2 Schoolbesturen formuleren meetbare doelen voor het beoordelen van de eigen leerlingprestaties.
- 3 Scholen met leerlingen met een leerlinggewicht meten de opbrengsten van de leerlingen in de groepen 1 en 2 door observaties of een vorm van toets of checklist.
- 4 Scholen bieden een gericht onderwijsaanbod voor cognitief talentvolle en hoogbegaafde leerlingen.
- 5 Scholen betrekken ouders actief bij de ontwikkeling en leerprestaties van hun leerlingen.
- 6 Er zijn geen zeer zwakke scholen meer.
- 7 Leraren volgen en analyseren systematisch de voortgang in de ontwikkeling van hun leerlingen. Ook stemmen leraren de instructie, de verwerking en de onderwijstijd af op verschillen in de ontwikkeling van leerlingen.
- 8 Leraren en schoolleiders werken voortdurend aan hun professionalisering en registreren zich als zodanig in het beroepsregister voor leraren en voor schoolleiders.
- 9 Scholen voeren een goed en effectief HRM-beleid. In dat verband is er voor leraren een bekwaamheidsdossier en vindt jaarlijks minimaal één formeel beoordelings- of functioneringsgesprek plaats. Daarnaast is de kwaliteit van de begeleiding en ondersteuning van beginnende leraren verbeterd.
- 10 Schoolbesturen geven hun onderwijsopbrengsten transparant weer via het te ontwikkelen informatiesysteem 'Vensters PO' en scholen gebruiken 'Vensters PO' voor het evalueren van de eigen opbrengsten en kwaliteit.

De prestatiebox is feitelijk een onderdeel van de lumpsum. Schoolbesturen kunnen zelf beslissen hoe zij de doelen van het bestuursakkoord willen realiseren. De prestatiebox bevat extra middelen die zijn bedoeld voor extra inspanningen ten aanzien van de beleidsambities van het po. Ook worden bestaande middelen opgenomen in de prestatiebox, zoals bestaande middelen voor taal en rekenen, middelen voor het programma 'Cultuureducatie met kwaliteit' en middelen voor het programma 'Wetenschap en techniek'. Daarnaast worden extra middelen beschikbaar gesteld voor experimenten met prestatiebeloning. Subsidiereregelingen komen met de introductie van de prestatiebox te vervallen.

Het totaalbedrag van de prestatiebox bedraagt 130 miljoen euro in 2012 en loopt op tot 150 miljoen euro vanaf 2014. Scholen in het po worden ondersteund vanuit hun sectorraad en kunnen bij de uitvoering van 'School aan zet' (zie verderop) desgewenst gebruikmaken van diens expertise en hulp. Voor opbrengstgericht werken en professionalisering van het (voortgezet) speciaal onderwijs is een bedrag beschikbaar dat oploopt tot 11,6 miljoen euro per jaar in 2016.

Het monitoren van de doelstellingen op het niveau van de hele sector vindt plaats via een *midterm-review* halverwege de looptijd van de akkoorden. Schoolbesturen verantwoorden zich via het jaarverslag over de inzet van middelen. Over de opbrengsten vindt horizontale verantwoording plaats via ‘Vensters PO’.

1.3 Doel

De prestatiebox is bedoeld om via een vereenvoudigde extra bekostiging landelijke doelen te realiseren wat betreft onderwijskwaliteit en onderwijsprestaties en om schoolbesturen binnen de gestelde beleidskaders en -doelen voldoende ruimte te bieden daarbij eigen keuzes te maken en accenten te leggen. Het uiteindelijke doel van de maatregel is de verbetering van de onderwijskwaliteit en -prestaties. Schoolbesturen worden geacht de extra middelen in te zetten voor de op sectorniveau afgesproken doelen. De doelen in het bestuursakkoord zijn in het kort: het versterken van de onderwijskwaliteit in de hele sector (onder meer het verminderen van het aantal zeer zwakke scholen en het aantal taal- en rekenzwakke scholen), meer opbrengstgericht werken, een beter aanbod voor begaafde leerlingen, professionalisering van leraren en schoolleiders en een beter HRM-beleid. Aangezien de beleidsambities aansluiten bij de reguliere kerntaken van de schoolbesturen in het po zullen ook de middelen uit de reguliere lumpsum moeten worden ingezet om deze te realiseren (Bestuursakkoord po 2012-2015).

1.4 Veronderstellingen over gedragsreacties en effecten

Een eerste veronderstelling lijkt te zijn dat zonder extra inspanningen de onderwijsprestaties in het primair onderwijs (verder) achteruit zullen gaan dan wel niet zullen verbeteren. Het bestuursakkoord in combinatie met de extra financiering via de prestatiebox moet schoolbesturen in het po committeren aan landelijke doelen op het gebied van onderwijsprestaties en onderwijskwaliteit. De wijze van sturing houdt het midden tussen enerzijds het generiek opleggen van landelijke doelen aan schoolbesturen en anderzijds het volledig overlaten aan besturen en scholen om autonoom de onderwijskwaliteit en prestaties te verbeteren. Het akkoord en de financieringswijze beogen recht te doen aan de bestaande rolverdeling in het onderwijs. De gedachte is dat de maatregel schoolbesturen voldoende ruimte biedt om hieraan per school een specifieke invulling te geven. Een andere veronderstelling is dat subsidies voor het onderwijsveld minder aantrekkelijk zijn (te rigide in de toepassing en omgeven door te veel bureaucratie) en dat de prestatiebox een gunstig alternatief is. Minder voor de hand liggend is dat alle besturen middelen krijgen via de prestatiebox, maar dat de hoogte van het bedrag niet gekoppeld is aan het aantal doelen of het ambitieniveau van scholen of besturen. Slechts het aantal leerlingen bepaalt de hoogte van het bedrag. Er vindt bovendien geen ‘afrekening’ plaats op school- of bestuursniveau als de doelen niet worden gehaald.

Effecten op instellingsniveau

In het po wordt de prestatiebox uitgekeerd aan de schoolbesturen. De verwachting is dat scholen als gevolg van de nieuwe lumpsumachtige financieringswijze wellicht minder, maar in elk geval niet meer bureaucratie zullen ervaren.

Het lijkt erop dat het akkoord een (inhoudelijk) sterker sturende rol van schoolbesturen vereist. De veronderstelling is dat schoolbestuurders hun directies zullen aansporen om de beleidsdoelen te implementeren op school- en klasniveau. Voor sommige besturen is dat een nieuwe invulling van hun rol (veel schoolbesturen in het po vervullen tegenwoordig nog een meer beheersmatige rol).

Een belangrijke aanname is verder dat scholen en leerkrachten (via hun bestuur of anderszins) op de hoogte zijn van het bestuursakkoord en ook weten dat ze extra middelen krijgen om zich in te zetten voor de doelen uit dit akkoord. Niet elk niveau in de schoolorganisatie hoeft volledig op de hoogte te zijn, maar de wetenschap dat er een akkoord en extra middelen zijn, kan bijdragen aan de betrokkenheid en het gevoel van urgentie bij de uitvoerenden.

De veronderstelling is dat op instellingsniveau een *bewustwordingsproces* op gang zal komen dat ertoe zal leiden dat de betrokkenen zich meer dan voorheen zullen willen inzetten voor het verbeteren van kwaliteit en prestaties. De extra financiering zal dus moeten bewerkstelligen dat het *draagvlak* voor de beleidsambities bij schoolbesturen en scholen toeneemt. Daarnaast wordt verondersteld dat de maatregel leidt tot een waarneembare *gedragsverandering* op school- en klasniveau.

Uiteindelijk moeten de veranderingen zichtbaar worden in verbeterde prestaties van leerlingen. Het is alleen de vraag hoeveel tijd hiervoor nodig is. Het is voorstelbaar dat scholen een aanlooptijd nodig hebben voordat de betere onderwijskwaliteit, opbrengstgericht werken, professionalisering, ouderbetrokkenheid en beter management doorwerken in de prestaties van leerlingen. Aan de andere kant: het is niet zo dat (alle) scholen op (al) deze punten 'vanaf nul' moeten starten.

Onbekend is wat het effect is van het min of meer sequentieel aan de verschillende doelen werken dan wel dat parallel doen. Het ligt minder in de verwachting dat scholen erin slagen om op alle doelen gelijktijdig heel gerichte actie te ondernemen.

Effecten op leerlingenniveau

De verwachting is dat leerlingen beter gaan presteren door de opbrengstgerichte werkwijze van hun leerkracht en diens professionalisering, waardoor hij of zij beter kan omgaan met verschillen. Dankzij een gericht onderwijsaanbod voor cognitief talentvolle en hoogbegaafde leerlingen zullen ook hun prestaties verbeteren. Daarnaast wordt een positief effect verwacht van een grotere ouderbetrokkenheid.

Effecten op leerkrachtniveau

Een van de verwachte effecten is dat leerkrachten opbrengstgerichter zullen gaan werken. Verder wordt verwacht dat zij zich zullen laten bijscholen om beter met verschillen tussen leerlingen te kunnen omgaan en dat zij in staat zullen zijn om begaafde leerlingen van een passend aanbod te voorzien. Dankzij een verbeterde HRM-kwaliteit

zouden leerkrachten zich meer of beter op hun plaats moeten gaan voelen in de schoolorganisatie. Maar er worden ook meer eisen aan leerkrachten gesteld, zodat ze beter gaan functioneren. Een goed HRM-beleid heeft naar verwachting ook effecten op een betere samenwerking tussen leerkrachten.

Onbedoelde effecten en beleidsafwijkingen

Een algemeen onbedoeld effect zou kunnen zijn dat sommige schoolbesturen of hun scholen zich niet gebonden voelen aan de afspraken die op sectorniveau zijn gemaakt. In theorie is het denkbaar dat besturen wel de middelen accepteren, maar geen enkele (substantiële) verandering (willen of kunnen) doorvoeren.

Een ander algemeen onbedoeld effect zou kunnen zijn dat de extra financiën voor het veld (de school, de leerkracht) niet als 'extra' (en aan bepaalde doelen verbonden) zichtbaar worden of door het veld niet als extra worden ervaren. Dit kan bij het veld het gevoel oproepen dat er meer eisen worden gesteld zonder dat daar iets tegenover staat (Passend onderwijs!), hetgeen drukt op de motivatie en uiteindelijk het succes van het beleid.

Een ander mogelijk risico betreft de hoeveelheid doelen uit het bestuursakkoord. Dit kan mogelijk contraproductief werken (gechargeerd gesteld: dit kan leiden tot scholen die trachten alles tegelijk uit te voeren, maar daardoor niets echt goed aanpakken of tot scholen die er vanwege de hoeveelheid niet eens aan beginnen).

Een van de onbedoelde effecten op instellingsniveau zou kunnen zijn dat scholen die al goed presteren zich niet uitgedaagd voelen om zich nog extra in te spannen. Een ander onbedoeld effect op instellingsniveau zou kunnen zijn dat minder goed presterende scholen zich juist niet extra gaan inspannen, maar rekenen op de betere prestaties van andere scholen. Het betreft hier immers doelrealisaties op sectorniveau en de afrekening (ook in financieel opzicht) vindt niet op schoolniveau plaats.

Weer een ander onbedoeld effect kan zich voordoen bij het doel voor begaafde leerlingen. Mogelijk zijn er scholen/leerkrachten/ouders die niet gecharmeerd zijn van het expliciet benoemen van verschillen tussen leerlingen in termen van cognitief talentvol en hoogbegaafd. In dat geval zou ook het aanbod voor deze categorie leerlingen niet goed van de grond komen.

1.5 Evaluatie

Deze maatregel bestaat uit een combinatie van een financieringswijze en op sectorniveau afgesproken doelen (het bestuursakkoord). In feite wordt hiermee vrij ongericht extra geld ingezet waarmee scholen kunnen gaan bijdragen aan het realiseren van doelen op sectorniveau. De doelen zijn gekwantificeerd op sectorniveau, maar verschillen in mate waarin ze concreet zijn en moeten bovendien vertaald worden naar schoolniveau. Enerzijds gaat het om outputdoelen: betere prestaties van talentvolle en begaafde leerlingen. Anderzijds gaat het om instrumentele doelen, die te maken hebben met het proces: opbrengstgericht werken, professionalisering, toename van de

ouderbetrokkenheid en verbeterd HRM-beleid. Het uiteindelijke doel is een verbetering van de prestaties en de onderwijskwaliteit (inclusief die van de leerkrachten en de schoolleiders) in het po. Hoewel dit geen doel is uit het bestuursakkoord stelt de maatregel tevens (nieuwe) eisen aan de schoolbesturen.

De centrale evaluatievragen zijn:

Wat betreft de opbrengstdoelen:

- 1 Treden na invoering van de prestatiebox veranderingen op in de prestaties en de onderwijskwaliteit in het primair onderwijs (bv. minder zwakke scholen of zwak presterende scholen op het gebied van taal en rekenen en betere prestaties van begaafde leerlingen)?
- 2 Zijn de veranderingen overtuigend toe te schrijven aan de maatregel?

Wat betreft de instrumentele doelen:

- 3 Treden er veranderingen op in de mate van opbrengstgericht werken, het betrekken van ouders bij de ontwikkeling en prestaties van hun kind, de mate van professionalisering en de kwaliteit van het HRM-beleid?
- 4 Is dit te danken aan de maatregel?

Wat betreft de sturingswijze:

- 5 Is deze vorm van financiering geschikt om de gestelde doelen te bereiken?

Effectevaluatie

De effectevaluatie is met name gericht op het kunnen vaststellen van het effect van het beleid op de opbrengstdoelen. Een effectevaluatie via een experiment of quasi-experiment is niet mogelijk. De maatregel wordt op hetzelfde moment voor alle po-scholen/besturen van kracht. Er zijn geen pilots gestart met het oog op de beleidsevaluatie.

De maatregel betreft een zeer globale sturingswijze op een scala aan doelen uit het bestuursakkoord, waarbij de gewenste uitkomsten op sectorniveau zichtbaar moeten zijn en waaraan individuele scholen bijdragen. De doelen (zowel procesgerelateerde als op opbrengst gerichte) die met deze maatregel worden nagestreefd, lopen zeer uiteen en interacteren bovendien met elkaar. Op dit moment is nog onduidelijk welke keuzes schoolbesturen met hun scholen zullen maken en hoe de doelen vervolgens op schoolniveau zullen worden geoperationaliseerd en geïmplementeerd. Verder wordt deze maatregel geïsoleerd genomen, maar zijn er meer beleidsinterventies gericht op het verbeteren van de onderwijskwaliteit en prestaties. Onduidelijk is ook of de maatregel op korte termijn tot grote waarneembare veranderingen zal leiden. Er is kortom sprake van een complexe relatie tussen eventuele resultaten en de beleidsinterventie, en er zijn vele factoren die de uitkomsten kunnen beïnvloeden.

Het is mogelijk om door middel van een voor- en nameting en eventueel tussen-tijdse metingen te monitoren of de prestaties zich na invoering van de prestatiebox

ontwikkelen in de gewenste richting. Het gaat dan om een stijging van de prestaties voor taal en rekenen bij alle leerlingen in het po en bij de 20% best presterende leerlingen in het po. Die betere prestaties moeten zichtbaar worden in hogere Cito-scores. Het is echter onduidelijk of eventueel betere prestaties dan aan de maatregel kunnen worden toegeschreven.

De instrumentele doelen, zoals verbeterde onderwijskwaliteit en professionalisering van schoolleiders en leerkrachten, zouden in de bevindingen van de Inspectie van het Onderwijs zichtbaar moeten worden. Maar ook hiervoor geldt dat een waargenomen verandering niet zonder meer aan de maatregel kan worden toegeschreven.

Verklarende evaluatie

De verklarende evaluatie is bedoeld om de mechanismen te achterhalen waardoor de maatregel wel, niet of niet in alle opzichten effect heeft. Het begrip ‘verklarende’ evaluatie is bij deze maatregel overigens niet dekkend, aangezien van een effectmeting geen sprake is. Vanwege de uniformiteit zullen we deze term toch verder gebruiken.

Het is van belang om vooraf de doelen en beleidsvragen van de verklarende evaluatie helder te krijgen. Deze doelen moeten worden uitgewerkt in duidelijke onderzoeksvragen. Vervolgens moet worden onderzocht welke gegevens nodig zijn om de betreffende vragen te kunnen beantwoorden, welke personen of actoren deze kunnen verschaffen en per wanneer ze beschikbaar zijn. Ook moet worden bepaald in welk stadium van het beleid de procesevaluatie moet plaatsvinden (HM Treasury 2011). De verklarende evaluatie kan niet bepalen of het beleid werkt (daarvoor is een effectevaluatie nodig), maar kan wel cruciale informatie opleveren over wat het beleid in de praktijk doet of juist niet en kan duidelijk maken of de beleidsveronderstellingen kloppen of niet.

Dit verklarende onderzoek naar de prestatiebox in het po is bedoeld om te achterhalen hoe de landelijke maatregel met sectoraal afgesproken doelen door scholen en leerkrachten wordt vertaald in de praktijk.

De vraag die in de verklarende evaluatie centraal staat, is welke veranderingen als gevolg van de maatregel op bestuurlijk, school-, klasniveau optreden en welke andere factoren de implementatie en uitvoering van de maatregel in de praktijk beïnvloeden in positieve dan wel negatieve zin. De veranderingen betreffen onder meer bewustwording, draagvlak en gedragsverandering.

Een eerste vraag is of scholen en leerkrachten weten dat er een bestuursakkoord is, of ze de inhoud ervan kennen en of ze weten dat er extra middelen voor deze doelen beschikbaar zijn gesteld. Daarna is het van belang duidelijk te krijgen hoe men concreet aan de slag is gegaan met de doelen uit het bestuursakkoord. Hierbij moet zicht worden verkregen op de rol van het bestuur, de schoolleiding en leerkrachten en ouders. Hoe is het beleid geland op scholen? In welke mate is er sprake van draagvlak voor de doelen? Zijn er verschillende prioriteiten gesteld, en waarom koos men daarvoor? Deze en andere vragen, die zijn afgeleid van bovengenoemde veronderstellingen per actor/niveau, moeten een plaats krijgen in het onderzoek.

Er zijn verschillende manieren om invulling aan de verklarende evaluatie te geven. Een goede mogelijkheid is een onderzoek met meerdere meetmomenten. Bij een kwantitatieve opzet zou het kunnen gaan om een vragenlijst onder schoolleiders en een onder leerkrachten. Bij een kwalitatieve opzet zou het om een selectie van cases (scholen) gaan, waarbij gestreefd wordt naar variatie in uitgangssituatie, prioriteiten in doelen en gekozen aanpak. Per uitgangssituatie zouden bij voorkeur enkele scholen moeten worden geselecteerd die ondanks een vergelijkbare uitgangssituatie hebben gekozen voor een verschillende aanpak.

In beide opzetten moeten in een eerste ronde (kort nadat de keuzes zijn gemaakt) de betreffende directies en leerkrachten worden bevraagd over hun motieven en verwachtingen ten aanzien van de gemaakte keuzes.

In een later stadium kan worden gevraagd naar de ervaringen, de geboekte vooruitgang (objectief en subjectief vastgesteld) en de tevredenheid met de gekozen aanpak. Doordat scholen in hun plannen moeten aangeven wat hun uitgangssituatie is en wat heel concreet hun ambitieniveau is, is het mogelijk vast te stellen welke vorderingen een school heeft gemaakt. Tevens kunnen ongewenste neveneffecten in kaart worden gebracht. Door een vergelijkende gevalsanalyse zou duidelijk moeten worden wat naar alle waarschijnlijkheid het beste werkt in welke omstandigheden.

In de verklarende evaluatie zou apart moeten worden ingegaan op de sturingswijze. Het is voor de beleidsmakers immers relevant om te weten hoe deze door het onderwijsveld wordt ervaren. Een onderzoek onder schoolbestuurders en schooldirecties kan daarover meer duidelijkheid geven. Hiertoe kan in een kwalitatief onderzoek, bijvoorbeeld met expertmeetings, worden gepeild wat de precieze voor- en nadelen van deze financieringswijze zijn, welke doelen men beter wel of niet in de prestatiebox kan onderbrengen, enzovoort.

Van een andere orde is de vraag naar de werking van de verantwoordingsplicht in een jaarverslag. Is met dit aspect in de ogen van de stakeholders afdoende geregeld dat de middelen op een verantwoorde wijze worden ingezet?

2 Verplichte eindtoets en verplicht leerlingvolgsysteem in het primair onderwijs

2.1 Achtergrond en aanleiding

Het kabinetsbeleid is gericht op het bevorderen van een ambitieuze en opbrengstgerichte leercultuur. Scholen moeten meer opbrengstgericht gaan werken, de prestaties van leerlingen moeten omhoog en er moet meer uit de beste leerlingen worden gehaald. Als onderdeel van dit beleid wordt in het primair onderwijs een centrale eindtoets ingevoerd en krijgen alle scholen de verplichting om een leerlingvolgsysteem te gebruiken. Met deze maatregelen geeft het kabinet gehoor aan onder andere de aanbevelingen van de commissie-Dijsselbloem, die van oordeel was dat alle leerlingen recht hebben op een objectieve vaststelling van hun kennis en vaardigheden aan het einde van de basisschool. Tot op heden ontbreekt het daaraan. De maatregel sluit in het basisonderwijs in belangrijke mate aan op de gangbare praktijk, waar de grote meerderheid van de scholen al meedoet aan de Cito-eindtoets en bijna alle scholen een variant van een leerlingvolgsysteem hanteren. In het speciaal basisonderwijs en het speciaal onderwijs is het gebruik van leerlingvolgsystemen en eindtoetsen nog in ontwikkeling.

2.2 Inhoud

Aanvankelijk zou met ingang van schooljaar 2012/13 in het basisonderwijs een centrale eindtoets worden ingevoerd. Het wetsvoorstel daartoe ligt momenteel echter nog in de Tweede Kamer en het is nog niet duidelijk wanneer de Kamer het zal behandelen. Als het parlement instemt met de maatregel, dan kan de toets waarschijnlijk op zijn vroegst in 2014 worden afgenomen. Het gebruik van een leerlingvolgsysteem is wel met ingang van schooljaar 2012/13 verplicht. Het speciaal basisonderwijs en het speciaal onderwijs volgen in 2015/16 (centrale eindtoets) en 2014/15 (verplicht leerlingvolgsysteem).

De centrale eindtoets omvat de onderdelen rekenen, taal en studievoordigheden en sluit nauw aan bij de Cito-eindtoets, die al op veel scholen (85%) in gebruik is. Ook het onderdeel wereldoriëntatie wordt aangeboden, maar dat is facultatief en telt niet mee bij de weging van de toetsresultaten. De toets zal op een later tijdstip in het schooljaar worden afgenomen dan nu het geval is (in het tweede deel van april in plaats van in februari). Voor leerlingen voor wie dat nodig is, is er een aangepaste versie (minder moeilijk) van de toets. Slechts in uitzonderlijke gevallen kunnen leerlingen worden vrijgesteld van deelname aan de eindtoets.

Doordat de verplichte eindtoets een voortzetting is van de Cito-eindtoets, brengt de invoering ervan voor veel scholen geen grote veranderingen met zich mee. Een verandering is wel dat scholen nog slechts bij hoge uitzondering leerlingen kunnen vrijstellen van deelname aan de eindtoets. Tot op heden gebeurt dat wellicht wat makkelijker, maar precieze informatie daarover ontbreekt. Ook nieuw zijn de verplaatsing naar een later tijdstip in het schooljaar en de stapsgewijze inbouw in de eerste jaren na de invoering

van de referentieniveaus voor taal en rekenen. De eindtoets laat dan zien of leerlingen aan het einde van het primair onderwijs voldoen aan het referentieniveau. Het Cito ontwikkelt de eindtoets, en de verantwoordelijkheid voor de afname komt te liggen bij het College van Examens (CvE). Dit is analoog aan de verdeling van verantwoordelijkheden bij de examens in het voortgezet onderwijs.

Naast de verplichte eindtoets wordt het gebruik van een leerlingvolgsysteem verplicht voor alle scholen. Veel basisscholen (95%) gebruiken overigens al een variant van een leerlingvolgsysteem, bijna altijd het systeem van het Cito. Scholen zijn vrij in de keuze van een leerlingvolgsysteem.

2.3 Doel

De invoering van de verplichte eindtoets en het verplichte leerlingvolgsysteem heeft verschillende doelstellingen:

- het bevorderen dat scholen opbrengstgericht werken en daardoor betere prestaties realiseren met hun leerlingen;
- het vergroten van de transparantie van leeropbrengsten op alle niveaus (voor leerlingen, voor groepen, voor scholen, voor besturen, op het niveau van de hele sector);
- het verbeteren van de aansluiting tussen het primair en het voortgezet onderwijs.

2.4 Veronderstellingen over gedragsreacties en effecten

De introductie van de centrale eindtoets en het verplichte leerlingvolgsysteem berust op verschillende veronderstellingen. Om leerlingen betere resultaten te laten halen is het belangrijk dat scholen opbrengstgericht gaan werken. Daarvoor is het nodig dat zij inzicht hebben in hun resultaten, dat ze deze evalueren en dat ze waar nodig hun doelen en hun werkwijze aanpassen. Het gebruik van toetsen is een belangrijke voorwaarde voor deze werkwijze. Het uiteindelijke doel is een verbetering van de leerprestaties. Opbrengstgericht werken is daarvoor een belangrijk middel. Dit vereist inzicht in de vorderingen van leerlingen, dat kan worden verschaft door de eindtoets en het leerlingvolgsysteem.

Verondersteld wordt ook dat een centrale eindtoets bijdraagt aan transparantie van leeropbrengsten op verschillende niveaus. Met de eindtoets komt eenduidige informatie beschikbaar over leeropbrengsten, zij het niet op alle gebieden. Voor individuele leerlingen en hun ouders wordt duidelijk hoe ze er aan het einde van het primair onderwijs voor staan, hoe ze zich verhouden tot andere leerlingen en hoe ze presteren ten opzichte van een absolute maatstaf (het referentieniveau). Scholen en besturen krijgen inzicht hoe hun resultaten zich verhouden tot die van andere scholen en besturen en tot het vastgestelde minimum (referentieniveau). Ook op stelselniveau verschaft de centrale eindtoets inzicht in de leeropbrengsten. Vergelijking van de resultaten door de jaren heen kan laten zien hoe de prestaties zich ontwikkelen op stelselniveau. Dankzij het verplichte karakter is er geen risico dat de eindtoets een vertekend beeld geeft doordat

niet alle scholen eraan meedoen of doordat een deel van de leerlingen de toets niet maakt. Transparantie van de gerealiseerde leeropbrengsten is ook een voorwaarde voor het kunnen afleggen van verantwoording aan de inspectie, ouders en andere stakeholders, en maakt beoordeling van het functioneren van scholen mogelijk.

De uitslag op de centrale eindtoets fungeert bij de overgang van het primair onderwijs naar het voortgezet onderwijs als tweede ‘objectief’ gegeven bij de plaatsing op verschillende niveaus, naast het schooladvies van de leerkracht. De toets draagt eraan bij dat leerlingen op een passend niveau van voortgezet onderwijs worden geplaatst. De huidige Cito-eindtoets vervult deze functie al voor veel leerlingen, maar na invoering van de verplichte eindtoets komt voor alle leerlingen een objectief en eenduidig gegeven beschikbaar. De veronderstelling is dat leerlingen daardoor nog meer dan voorheen op een passend niveau van voortgezet onderwijs kunnen worden geplaatst.

De uitgangspunten van de maatregel zijn op dit punt echter niet helemaal eenduidig. Met de verschuiving van de eindtoets naar een later moment in het schooljaar wordt namelijk tegelijkertijd beoogd het schooladvies meer gewicht te geven ten opzichte van de score op eindtoets. Hoewel het belangrijk wordt gevonden dat voor alle leerlingen een objectief gegeven beschikbaar komt, wordt de rol daarvan dus ook weer gereduceerd.

Omdat de referentieniveaus voor taal en rekenen aan de eindtoets zullen worden gekoppeld, draagt de verplichte eindtoets ook bij aan het realiseren van een doorlopende leerlijn bij de overgang van primair naar voortgezet onderwijs. Scholen voor voortgezet onderwijs weten hoe nieuwe leerlingen ervoor staan en kunnen daarop met hun onderwijs aansluiten (zie ook het evaluatieontwerp Introductie referentieniveaus in het primair, voortgezet en middelbaar beroepsonderwijs).

Effecten op schoolniveau

Voor scholen die nu al aan de Cito-eindtoets meedoen, brengt de verplichte eindtoets geen grote veranderingen met zich mee. Van belang is wel dat de eindtoets wordt verschoven naar een later tijdstip in het schooljaar. In het verleden is betoogd dat in het laatste schooljaar van de basisschool leertijd verloren gaat doordat de toets al in februari wordt afgenomen. Scholen zouden na de eindtoets de teugels laten vieren. Voor zover die veronderstelling klopt, zou het verlies aan leertijd moeten verminderen door de verschuiving naar een later tijdstip.

Voor scholen die nog niet aan de Cito-eindtoets deelnemen, betekent de verplichte eindtoets een grotere verandering. Het betreft bijvoorbeeld scholen die bezwaar maken tegen het te ‘smalle’ karakter van de eindtoets. Omdat de toets een rol speelt bij het transparant maken van leeropbrengsten, zullen deze scholen zich mogelijk gedwongen voelen er in hun onderwijs op te anticiperen. De toets brengt daardoor op deze scholen wellicht een accentverschuiving in het lesprogramma teweeg.

Een leerlingvolgsysteem is nu alleen verplicht voor zorgleerlingen. Volgens de inspectie beschikt 95% van de basisscholen echter al over een leerlingvolgsysteem, waarmee

veel scholen de vorderingen van alle leerlingen in kaart brengen. Bijna al deze scholen (97%) gebruiken het leerlingvolgsysteem van het Cito. Door de verplichting om een leerlingvolgsysteem te gebruiken verandert er dus niet zo veel. Het gebruik van het leerlingvolgsysteem als hulpmiddel voor opbrengstgericht werken laat overigens nog wel te wensen over, maar daarin brengt de maatregel geen verandering.

Effecten op leerlingenniveau

Voor leerlingen op scholen die tot op heden niet meedoen aan de Cito-eindtoets en voor leerlingen die om andere redenen niet meedoen aan de toets betekent de verplichte eindtoets een duidelijke verandering. Positief is dat zij nu volgens een eenduidige maatstaf inzicht krijgen in hoe ze er aan het einde van het primair onderwijs voor staan. Positief is ook dat de verplichte toets mogelijk leidt tot betere prestaties, omdat scholen vanwege de toets meer aandacht zullen besteden aan de kernvakken. Voor leerlingen die (zeer) zwak presteren, kan de verplichte deelname aan de eindtoets een belasting vormen, omdat de toets laat zien wat ze niet kunnen. De aangepaste versie van de toets voor zwakke leerlingen heeft het doel dat risico te verminderen.

Onbedoelde effecten en beleidsafwijkingen

Om het uiteindelijke doel – betere prestaties – te bereiken is meer nodig dan alleen het afnemen van een eindtoets en het hanteren van een leerlingvolgsysteem. Daartoe moet de informatie uit de toets en het leerlingvolgsysteem ook daadwerkelijk worden gebruikt als onderdeel van een opbrengstgerichte aanpak. Dit wordt in het beleid overigens onderkend: er worden ook andere maatregelen genomen om te bevorderen dat meer scholen opbrengstgericht gaan werken (de ‘Prestatiebox primair onderwijs’ en de operatie ‘School aan zet’, zie de desbetreffende evaluatieontwerpen).

Een (onbedoeld) neveneffect van de verplichte eindtoets kan zijn dat een versmalling optreedt van leeropbrengsten. De eindtoets richt zich slechts op een deel van de kern-doelen van het primair onderwijs. De functie die de toetsgegevens zullen vervullen (het transparant maken van verschillen tussen scholen en mogelijk ook het bevorderen van de concurrentie tussen scholen) brengt het risico met zich mee dat scholen hun onderwijsaanbod zullen inperken en minder aandacht zullen besteden aan doelen die niet in de toets aan bod komen.

De beoogde transparantie van de gerealiseerde leeropbrengsten kan eveneens onbedoelde effecten hebben. De toetsgegevens maken inzichtelijk wat het bereikte eindniveau is op een school. Voor scholen met een kansarme leerlingenpopulatie is het veel lastiger om een bepaald niveau te bereiken dan voor scholen met een kansrijke populatie. Als bij het transparant maken van de resultaten van scholen onvoldoende rekening wordt gehouden met dergelijke verschillen in uitgangssituatie, zal de vergelijking van resultaten van scholen tot verkeerde conclusies kunnen leiden. Ook kan de vergelijking dan als oneerlijk worden ervaren.

Scholen die vanwege hun hoge score op de eindtoets een naam hebben hoog te houden, zullen zich mogelijk minder soepel opstellen bij de toelating van in aanleg zwakke

leerlingen, die immers een ‘risico’ vormen voor de score op de eindtoets. Ook zou de doelstelling om ‘passend onderwijs’ voor zorgleerlingen aan te bieden binnen het reguliere onderwijs onder druk kunnen komen te staan als het zorgleerlingen betreft met een naar verwachting lage score op de eindtoets.

Een beoordeling van de leeropbrengsten die scholen realiseren in termen van leerwinst en toegevoegde waarde komt tegemoet aan deze bezwaren. Een dergelijke systematiek is echter nog niet operationeel. De inspectie houdt bij het beoordelen van leeropbrengsten op een school weliswaar rekening met de samenstelling van de leerlingenpopulatie, maar het bredere publiek zal wellicht minder genuanceerd omgaan met de informatie uit eindtoets en vooral letten op het bereikte eindniveau.

Een ander onbedoeld effect dat mogelijk kan optreden, heeft te maken met de verschuiving van de toets naar een later tijdstip in het jaar. Het doel daarvan is de invloed van de score op de eindtoets terug te dringen en het schooladvies van de basisschool een grotere rol te laten spelen bij de plaatsing in het voortgezet onderwijs dan nu het geval is. De eindtoets zou daardoor echter zijn corrigerende rol kunnen verliezen voor leerlingen die een ‘te laag’ of ‘te hoog’ advies krijgen.

Populaire scholen voor voortgezet onderwijs stellen nu soms extra hoge eisen aan het niveau van de score op de Cito-eindtoets. Het is de bedoeling dat die praktijk wordt teruggedrongen, maar het valt te bezien of dat daadwerkelijk lukt. Het staat scholen voor voortgezet onderwijs vrij om het toetsgegeven, dat nu pas later in het jaar beschikbaar komt, alsnog te hanteren als selectiemiddel.

2.5 Evaluatie

In het evaluatieontwerp staat de vraag centraal of met de verplichte eindtoets en het leerlingvolgsysteem de beoogde effecten worden gesorteerd (effectevaluatie) en welke processen en mechanismen verantwoordelijk zijn voor het al dan niet optreden van die effecten. Eerder werd door het Centraal Planbureau een ontwerp gemaakt voor de evaluatie van de invoering van de centrale eindtoets. De door het Centraal Planbureau voorgestelde effectevaluatie aan de hand van gegevens uit het Cool-cohortonderzoek bleek bij nader inzien niet te voldoen.¹ We presenteren hier een alternatieve aanpak.

Effectevaluatie

Om effecten vast te stellen is een vergelijking nodig tussen een groep die aan de interventie blootstaat en een vergelijkingsgroep die niet aan de interventie onderhevig is. Voor het samenstellen van een experimentele en een controlegroep kan gebruik worden gemaakt van het gegeven dat de verplichte eindtoets voortbouwt op de Cito-eindtoets, die nu al op 85% van de scholen wordt afgenomen. Voor de 85% scholen die al aan de toets deelnemen, verandert er niet veel en zij vormen dan ook de controlegroep die weinig of niet door het beleid wordt geraakt. De 15% scholen die nu nog niet aan de Cito-eindtoets meedoen, vormen de experimentele groep die met nieuw beleid te maken krijgt.

Omdat de verdeling over de experimentele en de controlegroep niet op toeval is gebaseerd, zullen de beide groepen van scholen in de uitgangssituatie mogelijk niet vergelijkbaar scoren op effectmaten. Het zou zo kunnen zijn dat leerlingen op scholen die niet meedoen aan de Cito-eindtoets minder vaardig zijn op het gebied van taal en rekenen dan vergelijkbare leerlingen op scholen die de toets wel afnemen. In het geval van een niet-gelijkwaardige experimentele en controlegroep kan een difference-in-differencebenadering worden toegepast. Daarbij worden op basis van een voor- en een nameting trends in de effectmaten bij beide groepen vastgesteld en met elkaar vergeleken. Een verschil in trend wijst op effecten. Essentieel hierbij is de veronderstelling dat de trend in beide groepen gelijk zou zijn als de maatregel niet zou zijn ingevoerd. Om de vergelijkbaarheid van de experimentele en de controlegroep te vergroten, kan de vergelijking worden toegespitst op scholen die op elkaar lijken wat betreft de samenstelling van de leerlingpopulatie en andere relevante kenmerken (bv. regio en mate van verstedelijking van de vestigingsplaats). Naarmate de beide groepen meer op elkaar lijken, is de aanname van gelijke trend aannemelijker.

Het effect van de introductie van het verplichte leerlingvolgsysteem is niet goed te meten. Bijna alle scholen hanteren namelijk al een leerlingvolgsysteem. Dit onderdeel van de maatregel brengt dus nauwelijks verandering voor scholen. Het gaat hier vooral om de manier waarop het leerlingvolgsysteem wordt gebruikt.

Welke effectmaten zijn beschikbaar voor de effecten die in de beleidstheorie worden verondersteld?

De inspectie beoordeelt of scholen *opbrengstgericht* werken. Om vast te stellen of de verplichte eindtoets een positief effect heeft op opbrengstgericht werken, kan een vergelijking worden gemaakt tussen scholen in de experimentele en de controlegroep. Als het verschil tussen beide groepen na invoering van de verplichte eindtoets afneemt, is dat een indicatie dat de maatregel effect heeft. Nader onderzocht moet worden of gegevens van de inspectie over opbrengstgerichtheid van een voldoende aantal scholen in de experimentele en de controlegroep beschikbaar zijn om deze vergelijking te kunnen maken.

Een duidelijke maatstaf voor *transparantie* ontbreekt. Overwogen zou kunnen worden om met een peiling onder betrokkenen na te gaan in hoeverre verschillende stakeholders (ouders, besturen, gemeenten) een beeld hebben van de resultaten die scholen in de experimentele en de controlegroep met hun leerlingen halen. In de beleidstheorie wordt verondersteld dat scholen in de experimentele groep in de ogen van betrokkenen op dat punt nu nog minder transparant zijn. Als blijkt dat na de invoering van de verplichte eindtoets de partijen die betrokken zijn bij de scholen in de experimentele groep een beter inzicht hebben gekregen in de bereikte leerresultaten, dan is dat een aanwijzing dat de maatregel het verwachte effect heeft. Voorwaarde daarbij is wel dat de trend bij de experimentele groep afwijkt van de trend bij de controlegroep.

Als maatstaf voor de aansluiting tussen het primair en het voortgezet onderwijs zou kunnen worden gekozen voor de mate waarin leerlingen in het voortgezet onderwijs een loopbaan volgen die in overeenstemming is met het door de basisschool afgegeven advies. Dit zou kunnen worden onderzocht op basis van gegevens uit BRON (Basisregistratie Onderwijsnummer). Als echter blijkt dat voor leerlingen in de experimentele groep het adviesgegeven niet beschikbaar is, moet een andere indicator worden gekozen. Te denken valt aan niveauwisselingen in de eerste jaren van de loopbaan in het voortgezet onderwijs. Deze zijn te beschouwen als een indicator voor een minder goede aansluiting tussen het primair en voortgezet onderwijs.

Voor de doelstelling van het verbeteren van *leerprestaties* zijn verschillende maatstaven denkbaar. Onderzocht kan worden of scholen in de experimentele groep gebruikmaken van een alternatieve eindtoets die kan worden getransformeerd naar de Cito-eindtoets die in de controlegroep wordt gebruikt. Mogelijk bieden ook de gebruikte leerlingvolgsystemen door scholen in de beide groepen een aanknopingspunt voor de vergelijking van leerprestaties. Een derde optie is dat leerlingen uit beide groepen van scholen worden gevolgd in hun loopbaan in het voortgezet onderwijs aan de hand van gegevens uit BRON. Indien de loopbanen van leerlingen uit de experimentele groep verbeteren ten opzichte van die in de controlegroep is dat een indicatie voor een effect van de eindtoets. Bij die vergelijking moet zo veel mogelijk rekening worden gehouden met relevante kenmerken over de achtergrond van leerlingen, zoals de etniciteit en de sociaaleconomische status.

Om loopbaangegevens uit BRON te kunnen gebruiken voor de effectevaluatie moet van leerlingen in het voortgezet onderwijs bekend zijn van welke basisschool ze komen.

Verklarende evaluatie

Een belangrijke vraag voor de verklarende evaluatie is hoe scholen het verplichte leerlingvolgsysteem en de eindtoets in de praktijk van het onderwijs gebruiken. Toetsgegevens zijn een voorwaarde voor opbrengstgericht werken, maar geen garantie. Als toetsgegevens niet worden benut om het onderwijs zo nodig bij te sturen, blijven de gewenste effecten op leerprestaties achterwege. In het verklarende onderzoek kan worden nagegaan hoe scholen omgaan met de informatie van de eindtoets en het leerlingvolgsysteem. Gebruiken scholen de gegevens inderdaad om opbrengstgericht onderwijs te organiseren?

Bij scholen die nu nog niet werken met de Cito-eindtoets kan in de verklarende evaluatie worden onderzocht hoe men aankijkt tegen de verplichte eindtoets. Roept de eindtoets weerstand op? Passen de scholen hun lesprogramma aan op de eindtoets? Welke voor- en nadelen ervaren de betrokkenen? Voldoet de aangepaste toets voor zwakkere leerlingen in de ogen van de betrokkenen? Ook is het van belang om te onderzoeken hoe scholen naar buiten toe omgaan met de resultaten uit de eindtoets. Gebruiken ze de toetsuitslag om hun resultaten zichtbaar te maken en zich te profileren ten opzichte van andere scholen? Of hangen ze de uitkomsten liever niet aan de grote klok?

Noot

- 1 Het aantal scholen in de experimentele groep – scholen die bij een eerdere meting niet en bij een latere meting wel meededen aan de Cito-eindtoets – was te gering voor een effectmeting.

3 Prestatiebox voortgezet onderwijs

3.1 Achtergrond en aanleiding

Zie paragraaf 1.1.

3.2 Inhoud

Heel algemeen gesteld is de prestatiebox een vorm van financiering bedoeld om de prestaties van leerlingen in het vo en de kwaliteit van het onderwijs en van de leraren en schoolleiders te verbeteren. De prestatiebox moet scholen en besturen stimuleren zich in te zetten voor landelijke beleidsdoelen. De prestatiebox is feitelijk een 'lege' regeling. Op sectorniveau zijn – in dit geval met vo-raad – meerjarige afspraken gemaakt over de doelen waarvoor de prestatiebox concreet zal worden ingezet gedurende de looptijd van het Actieplan, de periode 2012-2015. Het bestuursakkoord tussen de minister, de staatssecretaris en de vo-raad is op 14 december 2011 ondertekend. De middelen in de prestatiebox lopen op van 110 miljoen euro in 2012 naar 133 miljoen euro in 2015. De middelen worden toegekend op basis van het aantal leerlingen. Er is een algemeen bedrag beschikbaar plus een extra bedrag voor de 20% best presterende leerlingen in de bovenbouw van het vwo.

De prestatiebox wordt toegevoegd aan de lumpsummiddelen en biedt schoolbesturen de ruimte om zelf te bepalen wat de meest effectieve inzet van middelen is, passend bij hun schoolbeleid en hun eigen ambitie en uitgangssituatie. Scholen kunnen dus zelf bepalen hoe ze de middelen besteden om een of meer doelen uit het bestuursakkoord te realiseren. Zij kunnen besluiten bepaalde doelen niet te kiezen, maar moeten zich verantwoorden over alle doelen. De inspectie controleert niet hoe de middelen uit de prestatiebox worden ingezet, maar kijkt naar de prestaties van scholen/instellingen. Over de besteding van de middelen uit de prestatiebox leggen schoolbesturen wel verantwoording af via het jaarverslag. Verder moet de MR instemmen met de manier waarop de middelen worden ingezet.

De ambities uit het bestuursakkoord voor de periode 2012-2015 zijn de volgende:

- 1 Er is meer en gerichte aandacht voor de kernvakken Nederlands, Engels en wiskunde/rekenen, waardoor de leerprestaties van alle leerlingen op die vakken aantoonbaar verbeteren.
- 2 De leerprestaties van de 20% best presterende vwo-leerlingen stijgen aantoonbaar.
- 3 Scholen gaan meer opbrengstgericht werken, zodat in 2015 ten minste 50% van de scholen opbrengstgericht werkt.
- 4 Alle leraren zijn in staat het onderwijs af te stemmen op verschillen binnen de klas en kunnen opbrengstgericht werken, en alle leraren en schoolleiders voldoen aan de geldende bekwaamheidseisen en werken systematisch aan hun bekwaamheidsonderhoud (professionalisering).
- 5 Alle scholen voeren een goed en effectief HRM-beleid.

Het bestuursakkoord moet het commitment voor de beleidsdoelen creëren en is een eerste stap in de vertaalslag van de beleidsambities van het kabinet naar schoolniveau. Scholen moeten zelf de doelen operationaliseren op schoolniveau. De vo-raad biedt daarbij ondersteuning, en het ministerie van o c w doet dat ook via 'School aan zet' (zie verderop).

De doelen zijn gekwantificeerd en er zijn criteria, streefdoelen voor 2015 en tussentijdse doelen voor 2013 aan verbonden. De voortgang op de doelstellingen wordt jaarlijks op *sectorniveau* gemonitord. Er zijn drie ijkmomenten.

In 2012 geven alle scholen in het schoolplan aan wat ze met de prestatiebox gaan doen. Ze geven het o-niveau aan, de ambities (wat betreft het bestuursakkoord), de maatregelen en de inzet van middelen (zowel uit de prestatiebox als de reguliere middelen). Er vindt een landelijke steekproef van scholen plaats om het ambitieniveau te kunnen vaststellen.

In 2013 vindt na de zomer een midterm-review plaats, bedoeld om vast te stellen of de tussendoelen uit het bestuursakkoord op *sectorniveau* (en dus niet op het niveau van schoolbesturen of individuele scholen) zijn gehaald. Het resultaat wordt besproken met de vo-raad. Het is voor de overheid dan nog mogelijk om in te grijpen. Eventueel kan de regeling worden gewijzigd, bijvoorbeeld door de vrijblijvendheid weg te nemen.

In 2014/2015 vindt de eindevaluatie plaats over de voortgang van de sector ten opzichte van de streefsituatie. De prestatiebox is niet vrijblijvend. Als op *sectorniveau* te weinig voortgang wordt geboekt, dan wordt de inzet van middelen heroverwogen.

3.3 Doel

De prestatiebox beoogt via een vereenvoudigde extra bekostiging landelijke doelen te realiseren op het vlak van onderwijskwaliteit en onderwijsprestaties. Daarbij moet scholen binnen de gestelde beleidskaders en -doelen voldoende ruimte worden geboden om eigen keuzes te maken en accenten te leggen. Het uiteindelijk doel van de maatregel is de onderwijskwaliteit en onderwijsprestaties verbeteren. Scholen krijgen de middelen uit de prestatiebox boven op de lumpsumuitkering om zich in te zetten voor prestatiedoelen die in een bestuursakkoord op *sectorniveau* zijn vastgelegd. De doelen uit het bestuursakkoord zijn in het kort:

- betere prestaties in de kernvakken;
- stijgende prestaties van de 20% best presterende vwo-leerlingen;
- opbrengstgericht werken;
- professionalisering (waaronder het kunnen omgaan met verschillen);
- een effectief HRM-beleid.

3.4 Veronderstellingen over gedragsreacties en effecten

Een eerste veronderstelling lijkt te zijn dat zonder extra inspanningen de onderwijsprestaties in het voortgezet onderwijs (verder) achteruit zullen gaan dan wel niet

zullen verbeteren. Door landelijke doelen op het gebied van onderwijsprestaties en onderwijskwaliteit te concretiseren tot doelen op schoolniveau en daarvoor ook extra middelen beschikbaar te stellen (in de vorm van een prestatiebox) zullen scholen zich eerder willen en beter kunnen inzetten voor de doelrealisatie.

Landelijke doelen generiek opleggen zou inefficiënt zijn, omdat er zeer veel variatie tussen scholen bestaat in kwaliteit, prestaties en reeds eerder ingezette verbeteringen. De gedachte is dat de maatregel voldoende ruimte moet bieden om bij de situatie per school aan te sluiten. Voor scholen zijn desgewenst ondersteuningsmogelijkheden beschikbaar. Minder voor de hand liggend is dat alle scholen middelen krijgen via de prestatiebox, maar dat de hoogte van het bedrag niet gekoppeld is aan het aantal doelen of het ambitieniveau. Die financieringswijze moet bij scholen tot minder bureaucratie leiden dan bijvoorbeeld rond subsidies noodzakelijk is.

Effecten op instellingsniveau

Een belangrijke aanname is dat scholen en docenten op de hoogte zijn van het bestuursakkoord en ook weten dat ze extra middelen krijgen om zich meer te gaan inzetten voor de verwezenlijking van de doelen uit dit akkoord. Verondersteld wordt dat op schoolniveau een *bewustwordingsproces* op gang zal komen dat ertoe zal leiden dat de betrokkenen zich meer dan voorheen zullen willen inzetten voor het verbeteren van kwaliteit en prestaties. Door extra financiering aan de doelen te verbinden zal dus naar verwachting het *draagvlak* op scholen voor de beleidsambities toenemen.

Een andere verwachting is dat de afspraken op sectorniveau plus de extra financiering boven op de lumpsum zullen leiden tot een waarneembare *gedragsverandering* op school- en klasniveau. Dat wil zeggen: meer opbrengstgericht werken – zeker wat betreft de kernvakken – meer aandacht voor excellentie en hoogbegaafdheid bij leerlingen, verdere professionalisering van schoolleiding en docenten en een goed HRM-beleid. Het houdt in dat schoolbestuurders en schooldirectie zich actief inzetten om de plannen op schoolniveau te implementeren.

De verwachting is dat bovengenoemde gedragsveranderingen op scholen zichtbaar worden in verbeterde prestaties van leerlingen. Het is niet duidelijk hoeveel tijd daarvoor nodig is. Voorstelbaar is dat scholen een aanlooptijd nodig hebben voordat de betere onderwijskwaliteit en schoolorganisatie/-management doorwerken in de prestaties van leerlingen.

Ten slotte is de verwachting dat scholen als gevolg van de nieuwe lumpsumachtige financieringswijze wellicht minder, maar in elk geval niet meer bureaucratie zullen ervaren.

Effecten op leerlingenniveau

Leerlingen zullen zich door de toenemende aandacht op school voor prestaties in de kernvakken mogelijk meer voor deze vakken gaan inspannen. Door die toegenomen aandacht op school, een meer opbrengstgerichte werkwijze en professionalisering van docenten (waardoor zij beter met verschillen tussen leerlingen kunnen omgaan) zullen de prestaties van alle leerlingen in deze kernvakken verbeteren. Als scholen de middelen ook gebruiken voor een uitdagender en meer gedifferentieerd aanbod voor excellente leerlingen in het vwo, zullen bij deze leerlingen de prestaties ook in andere vakken

verbeteren.

Effecten op docentniveau

Een van de verwachte effecten is dat docenten opbrengstgerichter gaan werken, dat ze zich laten bijscholen om beter met verschillen tussen leerlingen te kunnen omgaan en dat ze de beter presterende vwo-leerlingen door bijvoorbeeld extra aanbod of verdieping nog beter kunnen laten presteren. Via een verbeterde HRM-kwaliteit zouden docenten zich bovendien meer of beter op hun plaats moeten voelen in de schoolorganisatie. Daarentegen zullen er meer dan in het verleden eisen worden gesteld aan de kwaliteiten van docenten. Een goed HRM-beleid zal naar verwachting ook leiden tot een betere samenwerking tussen docenten.

Onbedoelde effecten en beleidsafwijkingen

Een algemeen onbedoeld effect zou kunnen zijn dat sommige scholen zich niet gebonden voelen aan de afspraken die op sectorniveau zijn gemaakt. In theorie is het denkbaar dat scholen wel de middelen accepteren, maar geen enkele verandering doorvoeren.

Een ander algemeen onbedoeld effect zou kunnen zijn dat de extra financiën voor het veld (de school, de docent) niet als 'extra' (en aan bepaalde doelen verbonden) zichtbaar worden of door het veld niet als extra worden ervaren. Dit kan bij het veld het gevoel oproepen dat er meer eisen worden gesteld zonder dat daar iets tegenover staat.

Een van de onbedoelde effecten op instellingsniveau zou kunnen zijn dat scholen die toch al goed presteren zich niet uitgedaagd voelen om zich nog extra in te spannen. Weliswaar staat in het bestuursakkoord dat scholen de kwalificatie *excellent* kunnen verdienen, maar dit kan op zoveel aspecten worden toegekend (cognitief, maar bijvoorbeeld ook: ouderbetrokkenheid) dat hiervan naar verwachting nauwelijks een prikkel uitgaat om zich daarop te willen onderscheiden. Een ander onbedoeld effect op instellingsniveau zou kunnen zijn dat minder goed presterende scholen meevaren op de betere prestaties van andere scholen. De afrekening (ook financieel) vindt immers niet op schoolniveau plaats.

Verder zouden instellingen door zich met name te focussen op de kernvakken andere vakken kunnen gaan 'verwaarlozen'. Dat zou kunnen resulteren in een onbedoeld effect op leerling-niveau. Als leerlingen deze kernvakken minder interessant vinden dan andere vakken, dan daalt hun motivatie bij een te eenzijdige focus daarop. Dat geldt wellicht vooral voor minder cognitief ingestelde leerlingen (bv. in het vmbo). Omdat het hier gaat om doorstroomrelevante vakken is het tevens denkbaar dat leerlingen een louter instrumentele houding aannemen en geen intrinsieke motivatie voor deze vakken ontwikkelen.

Als leerlingen ten slotte weten of ze wel of niet tot de 20% beste vwo-leerlingen behoren, zou dat ten koste van de sfeer kunnen gaan. Zeker omdat het een vrij willekeurig criterium betreft, dat niet is gebaseerd op een absoluut kennisniveau.

3.5 Evaluatie

Deze maatregel bestaat uit een combinatie van een financieringswijze en op sector-niveau afgesproken doelen. In feite wordt hiermee vrij ongericht extra geld ingezet waarmee scholen kunnen gaan bijdragen aan het realiseren van doelen op sectorniveau. De doelen zijn weliswaar gekwantificeerd, maar lopen sterk uiteen en kunnen per school geen of een heel verschillende invulling krijgen. Enerzijds gaat het om outputdoelen: betere prestaties in kernvakken en betere prestaties van de beste vwo-leerlingen. Anderzijds gaat het om instrumentele doelen, die te maken hebben met het proces: opbrengstgericht werken, professionalisering, verbetering van het HRM-beleid, enzovoort. Het uiteindelijke doel is een verbetering van de prestaties en de onderwijskwaliteit (ook van de leraren en de directies) in het vo.

De centrale evaluatievragen zijn:

Wat betreft de opbrengstdoelen:

- 6 Treden na invoering van de prestatiebox veranderingen op in de prestaties en de onderwijskwaliteit in het voortgezet onderwijs?
- 7 Zijn de veranderingen overtuigend toe te schrijven aan de maatregel?

Wat betreft de instrumentele doelen:

- 8 Treden er veranderingen op in de mate van opbrengstgericht werken, de mate van professionalisering en de kwaliteit van het HRM-beleid?
- 9 Is dit te danken aan de maatregel?

Wat betreft de sturingswijze:

- 10 Is deze vorm van financiering geschikt om de gestelde doelen te bereiken?

Effectevaluatie

Een effectevaluatie via een experiment of in een quasi-experimentele opzet is niet mogelijk. De maatregel wordt op hetzelfde moment voor alle vo-scholen van kracht. Er zijn geen pilots gestart met het oog op de beleidsevaluatie.

De maatregel betreft een zeer globale sturingswijze op een scala aan doelen uit het bestuursakkoord, waarbij de gewenste uitkomsten op sectorniveau zichtbaar moeten zijn en waaraan individuele scholen bijdragen. De doelen (zowel procesgerelateerde als op opbrengstgerichte) die met deze maatregel worden nagestreefd, lopen zeer uiteen en interacteren bovendien met elkaar. Op dit moment is nog onduidelijk welke keuzes scholen zullen maken en hoe de doelen vervolgens op schoolniveau zullen worden geoperationaliseerd. Verder wordt deze maatregel niet geïsoleerd genomen, maar zijn er meer beleidsinterventies gericht op het verbeteren van de onderwijskwaliteit en prestaties. Ten slotte kan niet verwacht worden dat de maatregel op heel korte termijn tot grote waarneembare veranderingen zal leiden. Er is kortom sprake van een complexe relatie tussen eventuele resultaten en de beleidsinterventie, en er zijn vele factoren die de uitkomsten kunnen beïnvloeden.

Het is mogelijk om door middel van een voor- en nameting en eventueel tussentijdse metingen te monitoren of de prestaties in de kernvakken zich na invoering van de prestatiebox ontwikkelen in de gewenste richting. Het betreft betere prestaties in de kernvakken Nederlands, Engels en wiskunde/rekenen bij alle leerlingen in het voortgezet onderwijs en aantoonbaar betere prestaties van de 20% best presterende vwo-leerlingen. Die betere prestaties moeten ook zichtbaar worden in de PISA-resultaten. Het is dan echter onduidelijk of eventueel betere prestaties wel aan de maatregel kunnen worden toegeschreven.

Instrumentele doelen als verbeterde onderwijskwaliteit en professionalisering van schoolleiders en docenten zouden in de bevindingen van de inspectie zichtbaar moeten zijn. Maar ook hiervoor geldt dat een waargenomen verandering niet zonder meer aan de maatregel kan worden toegeschreven.

Verklarende evaluatie

De belangrijkste opdracht voor een verklarende evaluatie is te achterhalen hoe de landelijke maatregel met sectoraal afgesproken doelen door scholen wordt vertaald. Welke veranderingen zijn als gevolg van de prestatiebox en het bestuursakkoord opgetreden op schoolniveau, op docent- en op leerlingenniveau? Ook nu geldt weer dat aangezien er geen effectmeting plaatsvindt de term ‘verklarende’ evaluatie niet helemaal op zijn plaats is, maar hier vanwege de uniformiteit toch wordt gebruikt.

In deze ‘verklarende’ evaluatie is een eerste belangrijke vraag of scholen en docenten weten dat er een bestuursakkoord is, of ze de inhoud ervan kennen en of ze weten dat er extra middelen voor beschikbaar zijn gesteld. Vervolgens is het van belang duidelijk te krijgen of en hoe scholen aan de slag zijn gegaan met de doelen uit het bestuursakkoord. Een onderzoek met meerdere meetmomenten is zinvol. Dat zou zowel een meer kwantitatief als een meer kwalitatief onderzoek kunnen zijn. In het ene geval betreft het een survey onder docenten en directies; in het andere geval een diepteonderzoek onder een selectie van cases (scholen), waarbij gestreefd wordt naar variatie in uitgangssituatie, gekozen doelen en aanpak. Per uitgangssituatie zouden bij voorkeur enkele scholen moeten worden geselecteerd die ondanks een vergelijkbare uitgangssituatie hebben gekozen voor een verschillende aanpak.

In een eerste ronde (kort nadat de keuzes zijn gemaakt) moeten de betreffende directies en docenten worden bevraagd over hun motieven en verwachtingen ten aanzien van de gemaakte keuzes. In een later stadium kan worden gevraagd naar de ervaringen, de geboekte vooruitgang (objectief en subjectief vastgesteld) en de tevredenheid met de gekozen aanpak. Doordat scholen in hun plannen moeten aangeven wat hun uitgangssituatie is en wat heel concreet hun ambitieniveau is, is het mogelijk om vast te stellen welke vorderingen een school heeft gemaakt. Tevens kunnen ongewenste neveneffecten in kaart worden gebracht. Door een vergelijkende gevalsanalyse zou duidelijk moeten worden wat het beste werkt in welke omstandigheden.

Een andere vraag die een plaats zou moeten krijgen in de verklarende evaluatie is die naar de rol van de MR. Is de MR om instemming gevraagd bij het inzetten van middelen? Zijn de verschillende geledingen van de MR van mening dat zij daarop invloed konden uitoefenen?

4 Aanscherping exameneisen in het voortgezet onderwijs

4.1 Achtergrond en aanleiding

Er is twijfel ontstaan over de waarde van het diploma in het voortgezet onderwijs. Leerlingen kunnen een diploma halen terwijl ze gemiddeld een onvoldoende scoren op het centraal examen. Dat onvoldoende cijfer kunnen ze namelijk compenseren met een hoger gemiddeld cijfer voor het schoolexamen. De mogelijkheid dat leerlingen een diploma kunnen halen met een gemiddeld onvoldoende centraal examen wordt beschouwd als een 'weeffout' in het systeem die moet worden hersteld. Het diploma van leerlingen met een onvoldoende cijfer voor het centraal schriftelijk heeft niet het gewenste niveau. Daardoor komen deze leerlingen mogelijk in problemen in het vervolgonderwijs. Ook blijkt dat de kloof tussen het gemiddelde cijfer voor het schoolexamen en het gemiddelde cijfer voor het centraal examen in de loop der jaren is toegenomen. Een grote kloof tussen die twee cijfers tast het vertrouwen in diploma's aan. De aanscherping van de exameneisen maakt deel uit van het bredere beleid gericht op versterking van de kwaliteit van het voortgezet onderwijs: introductie van referentieniveaus voor taal en rekenen/wiskunde, het handhaven van de urennorm en investeren in de kwaliteit van leraren.

In dit evaluatieontwerp richten we ons op de aanscherping van de exameneisen en niet op het terugdringen van de kloof tussen het cijfer voor het schoolexamen en het centraal examen. De aanscherping van de exameneisen kan ook invloed hebben op de kloof tussen de beide examens, maar voor het terugdringen van deze kloof is er een aparte beleidslijn.

4.2 Inhoud

Met ingang van 2011/2012 worden de eindexameneisen in het voortgezet onderwijs in twee stappen verzwaaard. Bij het eindexamen van 2011/2012 zal het niet meer mogelijk zijn om een gemiddeld onvoldoende cijfer voor het centraal eindexamen te compenseren met hogere cijfers voor het schoolexamen. Het gemiddelde cijfer voor het centrale examen moet voldoende zijn. Tegelijkertijd wordt in het vmbo in de basisberoepsgerichte leerweg de gewichtverhouding tussen het schoolexamen en het centraal eindexamen gewijzigd. Tot 2010/2011 telde het schoolexamen bij de berekening van het gemiddelde eindcijfer daar dubbel zo zwaar als het cijfer voor het centraal examen. Met ingang van 2011/2012 tellen beide onderdelen even zwaar, net als in de andere vmbo-leerwegen en in havo en vwo.

In 2012/2013 wordt in havo en vwo een tweede stap gezet met de zogenoemde kernvakken. Leerlingen mogen dan niet meer dan één vijf halen voor de kernvakken Nederlands, Engels en wiskunde (dit betreft het gemiddelde van school- en centraal examen).

Om de kloof tussen het gemiddelde cijfer voor school- en centraal schriftelijk examen terug te dringen is vanaf 2013 een ‘escalatieladder’ van maatregelen van kracht. De uiterste sanctie is het ontnemen van de eindexamenlicentie van scholen waar de kloof tussen het schoolexamen en het centraal eindexamen meerdere jaren achtereen te groot is. Scholen met een grote kloof tussen beide examens krijgen hulp aangeboden van de Stichting voor de Leerplanontwikkeling en de vo-raad om de kwaliteit van hun school-examens te verbeteren.

4.3 Doel

Het doel van de maatregel is het borgen van de kwaliteit van het voortgezetonderwijs-diploma en het behoud c.q. herstel van het vertrouwen in het niveau van de gediplomeerden. Omdat de exameneisen strenger worden, gaat het niveau van gediplomeerden omhoog.

De kernvakkeneis in havo en vwo richt zich met name ook op het beter voorbereiden van gediplomeerden op een vervolgstudie in het hoger onderwijs en – in het verlengde daarvan – op een beter verloop van de studieloopbaan in het hoger onderwijs.

4.4 Veronderstellingen over gedragsreacties en effecten

Een centrale *veronderstelling* is dat de mogelijkheid om met onvoldoende cijfers voor het centrale examen een diploma te kunnen halen de waarde van het voortgezet-onderwijsdiploma aantast. Het centraal examen biedt een betere kwaliteitsgarantie dan het schoolexamen, omdat ‘het bestaat uit kwalitatief hoogwaardige, landelijk genormeerde toetsen’. Leerlingen met een gemiddeld onvoldoende cijfer voor het centraal examen hebben een te laag niveau. Op scholen met een grote kloof tussen het schoolexamen en het centraal examen is de waarde van het diploma onvoldoende geborgd. Een andere veronderstelling is dat de kernvakken Nederlands, wiskunde en Engels belangrijker zijn voor het succes in het hoger onderwijs dan andere vakken.

Effecten op leerlingenniveau

Door invoering van de nieuwe exameneisen neemt het niveau van de gediplomeerden per definitie toe. Het realiseren van deze doelstelling zit in de maatregel ‘ingebakken’. Een belangrijke veronderstelling daarbij is dat het niveau van het centrale examen in opeenvolgende jaren gewaarborgd blijft, ook als dat tot gevolg zou hebben dat veel meer leerlingen zullen zakken.

De strengere exameneisen kunnen verschillende gedragsreacties van leerlingen uitlokken. Leerlingen kunnen beter hun best gaan doen, maar ze kunnen ook voorzichtiger worden met hun keuzes in termen van het niveau en/of het profiel (kiezen voor een lager niveau, kiezen voor een minder moeilijk profiel, vaker kiezen voor de minder zware wiskundevariant vanwege de kernvakkeneis).

De verwachting is dat gediplomeerden dankzij de strengere eisen beter zijn toegerust voor het vervolgonderwijs. Dit zou zichtbaar moeten worden in een beter verloop van hun studieloopbanen in het hoger onderwijs.

Effecten op schoolniveau

Scholen kunnen proberen de kwaliteit van hun onderwijs te verbeteren, met name in de hogere jaren. Ze kunnen daarnaast ook strenger en selectiever worden bij de toelating van leerlingen tot de verschillende niveaus en/of strengere overgangsnormen hanteren om de slagingskansen op peil te houden.

De nieuwe eisen kunnen uiteenlopende effecten hebben op de examenresultaten. Simulatie van de nieuwe eisen op het laatste cohort dat examen heeft gedaan volgens de oude regels laat zien dat het slagingspercentage sterk kan afnemen (zie de simulatie van vo-raad). Of dat werkelijk gebeurt, is afhankelijk van de gedragsreacties bij scholen en leerlingen. De optimale uitkomst van de maatregel is dat het aantal examenkandidaten en het slagingspercentage op de verschillende niveaus op peil blijven dankzij meer inspanning van zowel leerlingen als scholen. In een ander scenario blijven de slagingspercentages op peil, maar neemt het aantal kandidaten op de hogere niveaus af als gevolg van voorzichtiger keuzes en/of strengere toelatingseisen en overgangsnormen. Als verschillende gedragsreacties zich gelijktijdig voordoen (zowel meer inspanning van leerlingen en scholen als voorzichtiger keuzes en strengere toelatings-/overgangsnormen), zal het resultaat tussen deze uitersten in zitten: het slagingspercentage neemt dan af – maar niet zo sterk als de simulaties suggereren – en er zijn minder kandidaten op de hogere niveaus van voortgezet onderwijs. Door ander beleid – meer nadruk op de basisvaardigheden en introductie van de referentieniveaus voor taal en rekenen – kunnen de gevolgen voor de slagingspercentages of voor het aantal leerlingen dat deelneemt aan de hogere niveaus van voortgezet onderwijs meevallen.

De kernvakkeneis in de havo en het vwo richt zich op drie vakken. Een mogelijk effect is dat de cijfers voor deze vakken verbeteren ten opzichte van die voor andere vakken, vanwege de negatieve gevolgen van een onvoldoende cijfer voor de kernvakken.

De nieuwe exameneisen richten zich als zodanig niet op de kloof tussen het school-examen en het centrale examen. De strengere exameneisen kunnen wel effect hebben op deze kloof. Omdat de leerlingen niet meer kunnen slagen met een gemiddeld onvoldoende cijfer voor het centraal examen, zullen de zwakste leerlingen beter moeten presteren bij het centrale examen. De kloof tussen het schoolexamen en het centrale examen zal doordoor mogelijk verminderen.

Onbedoelde effecten en beleidsafwijkingen

Enkele mogelijke onbedoelde effecten zijn hiervoor al aangestipt. Scholen worden wellicht voorzichtiger met de toelating van leerlingen tot hogere niveaus. Als ze dat te rigouzeus doen, is er het risico dat talent onbenut blijft. Daarnaast bestaat het risico dat leerlingen vooral veilige keuzes gaan maken, waardoor ze te laag inzetten.

4.5 Evaluatie

Hoe kan worden vastgesteld of de beoogde effecten worden bereikt? De centrale doelstelling van een betere waarborg van het niveau van diploma's en van gediplomeerden zit 'ingebakken' in de maatregel. Het middel (strengere exameneisen) en de doelstelling (een beter borging van het diploma en een hoger kennisniveau van gediplomeerden) vallen samen.

Toch is het belangrijk om na te gaan hoe de maatregel in de praktijk uitwerkt. Welke gedragsreacties roept de maatregel op bij leerlingen en scholen? Gaan leerlingen beter hun best doen of maken ze voorzichtiger keuzes om minder risico te lopen? Voeren scholen hun inspanningen op of gaan ze strengere regels hanteren rond toelating en zittenblijven? Het monitoren van veranderingen in slagingspercentages, in aantallen examenkandidaten op de verschillende niveaus, in de verdeling over profielen in de bovenbouw en in zittenblijfpercentages in eerdere jaren geeft aanwijzingen voor het optreden van de verschillende gedragsreacties.

Effectevaluatie

De maatregel wordt integraal ingevoerd en is van toepassing op alle scholen. Het ideaal van een random samengestelde experimentele groep die aan de maatregel wordt blootgesteld en een niet aan het beleid onderworpen controlegroep is dus niet haalbaar. Wel is het mogelijk om opeenvolgende cohorten – voor en na de invoering van de strengere exameneisen – met elkaar te vergelijken. Als dat gebeurt op het niveau van het totale voortgezet onderwijs is het echter onzeker of de waargenomen veranderingen in slagingspercentages, aantallen examenkandidaten, enzovoort zijn toe te schrijven aan de nieuwe exameneisen. Ook andere beleidslijnen, zoals de invoering van de referentieniveaus voor taal en rekenen of meer nadruk op opbrengstgericht werken, kunnen dan van invloed zijn op de waargenomen veranderingen. Voor het monitoren van de doelstellingen van het beleid is dit wellicht geen probleem, maar wel voor het doen van uitspraken over effecten.

Een alternatieve aanpak van de effectevaluatie gaat uit van groepen leerlingen die van elkaar verschillen in de mate waarin de nieuwe exameneisen gevolgen voor hen hebben. Er zullen waarschijnlijk groepen zijn die al in hoge mate voldoen aan de nieuwe exameneisen en voor wie de maatregel daarom weinig impact heeft. Daarnaast zullen er groepen zijn voor wie de maatregel veel meer impact heeft, omdat ze veel minder voldoen aan de nieuwe strengere eisen. Beide soorten groepen verschillen van elkaar in de mate waarin ze zijn blootgesteld aan het beleid en worden geprikkeld door de nieuwe eisen. De groep waarvoor het beleid veel impact heeft, fungeert als de experimentele groep; de groep waarvoor het beleid weinig impact heeft, is de controlegroep.

Omdat de zo samengestelde experimentele en controlegroep niet vergelijkbaar zijn, moet de effectmeting worden gebaseerd op de difference-in-differencebenadering. Daarbij worden aan de hand van een voor- en een nameting de trends in beide groepen vergeleken. Bij de trendvergelijking kan rekening worden gehouden met de stapsgewijze

invoering van de aanscherping van de eisen in havo en vwo (eerst een gemiddeld voldoende cijfer voor het centraal examen, in het volgend jaar ook de kernvakkeneis). Onder aanname dat de trend zonder de nieuwe maatregel in beide groepen gelijk zou zijn, kan uit een verschil in trend bij beide groepen het effect van de maatregel worden afgeleid.

De experimentele en de controlegroepen kunnen worden opgespoord door de nieuwe examenregeling te simuleren op cohorten die examen hebben gedaan onder de oude regels. Bij het zoeken naar groepen zou kunnen worden gekeken naar *profielen* – hebben de nieuwe regels groter gevolgen voor leerlingen in bepaalde profielen? Volgens de simulatie van de vo-raad is er in ieder geval een verschil tussen jongens en meisjes. Op basis van huidige situatie concludeert de vo-raad dat de nieuwe normen grotere gevolgen hebben voor de slaagcijfers van meisjes dan van jongens. De oorzaak is dat meisjes vaker dan jongens een onvoldoende centraal examen compenseren met een hoger cijfer voor het schoolexamen. Daarnaast is er wellicht een verschil naar herkomst van de leerling (allochtoon versus autochtoon). Ook kan worden gekeken naar scholen en leerlingen in verschillende regio's.

De effecten van strengere exameneisen kunnen worden afgemeten aan maatstaven zoals slagingspercentages, aantallen examenkandidaten, de verdeling over niveaus en profielen in het voortgezet onderwijs, de zittenblijfpercentages en de afstroom naar lagere niveaus in de jaren voor het examen. Deze gegevens zijn beschikbaar via BRON (Basisregistratie Onderwijsnummer). In het kader van het monitoren van veranderingen rondom de invoering van de nieuwe eindexameneisen worden al veel van deze gegevens verzameld. De effectevaluatie kan hierbij aansluiten. Een aparte gegevensverzameling lijkt dus niet nodig, een uitsplitsing van de gegevens langs de hiervoor geschetste lijnen volstaat.

Vergelijking van loopbanen in het hoger onderwijs

De strengere exameneisen hebben ook tot doel om leerlingen beter voor te bereiden op het vervolgonderwijs. Onderzocht zou kunnen worden of leerlingen het dankzij de strengere exameneisen beter zullen doen in het vervolgonderwijs. Een vergelijking tussen de loopbanen in het vervolgonderwijs van opeenvolgende instroomcohorten voldoet hiervoor niet. Er verandert de komende tijd veel tegelijk in het middelbaar beroepsonderwijs en het hoger onderwijs, en veranderingen in studieloopbanen die uit een vergelijking van opeenvolgende instroomcohorten naar voren komen zijn daardoor niet aan een bepaalde oorzaak toe te schrijven.

Wel zou binnen *hetzelfde* instroomcohort in het vervolgonderwijs de loopbaan kunnen worden vergeleken van groepen die zoveel mogelijk op elkaar lijken maar van elkaar verschillen wat betreft het wel of niet voldoen aan de nieuwe exameneisen. De groep die voldeed aan de oude eisen, maar niet aan de nieuwe eisen wordt dan vergeleken met de groep die aan de nieuwe eisen voldoet. Hier kan bij studenten met een havo/vwo-achtergrond nog onderscheid worden gemaakt tussen de eis dat het cijfer voor het centraal schriftelijk voldoende moet zijn en de kernvakkeneis.

Een andere optie is dat wordt gebruikgemaakt van het gegeven dat een deel van de studenten pas na een onderbreking van een of twee jaar begint aan het vervolgonderwijs. Als de eerste lichter studenten die een vmbo, havo of vwo-diploma hebben gehaald onder de nieuwe regeling begint aan het vervolgonderwijs, zullen hierdoor ook nog studenten aan de opleiding beginnen die een diploma hebben gehaald onder de oude regeling. Hierdoor is het mogelijk om binnen hetzelfde instroomcohort studenten gediplomeerd onder het nieuwe en het oude regiem in hun loopbaan met elkaar te vergelijken.

De effectmaten hebben betrekking op de studievoortgang (snelheid, diplomarendement en eventueel cijfers). Gegevens voor deze vergelijking zijn te vinden in de hogeronderwijscohorten waarover het CBS publiceert.

Verklarende evaluatie

Bij de verklarende evaluatie staan de gedragsreacties van scholen en leerlingen centraal. Onderzoek onder schoolleiders kan inzicht geven in de maatregelen die op schoolniveau zijn genomen om in te spelen op de nieuwe, strengere slaagnormen (bv. selectiebeleid, bevorderingsbeleid, intensivering van de examenvoorbereiding, meer aandacht voor de kernvakken of maatregelen rond de schoolexamens). Hiertoe kan een gestandaardiseerde (kwantitatieve) aanpak worden gehanteerd, maar ook een meer kwalitatieve aanpak (interviews) of een combinatie van beide.

Een relevante vraag voor onderzoek onder leerlingen is welke rol de strengere exameneisen hebben gespeeld bij hun keuze voor een opleiding van een bepaald niveau of voor een bepaald profiel. Hebben de leerlingen hun keuze aangepast met het oog op de strengere exameneisen of hebben ze besloten harder te werken voor school? Een onderzoek naar de reacties van leerlingen zou zich moeten richten op de groep leerlingen die in de loop van het voortgezet onderwijs met de nieuwe eisen is geconfronteerd.

5 Prestatiebox middelbaar beroepsonderwijs

5.1 Achtergrond en aanleiding

Zie paragraaf 1.1.

5.2 Inhoud

In het mbo wordt net als in het po en het vo een prestatiebox ingevoerd om de doelen uit de actieplannen van het kabinet te realiseren. Die doelen zijn: het verbeteren van de kwaliteit van het onderwijs, het management, de docenten en de gekwalificeerde uitstroom en het vergroten van de tevredenheid van het bedrijfsleven. De maatregel komt in de plaats van allerlei aparte subsidieregelingen voor specifieke doelen en is een aanvulling op de reguliere lumpsum. Wel is er een verschil met de andere sectoren (naast de beoogde doelen). In het mbo bestaat de prestatiebox uit een vast bedrag en een variabel bedrag. Het vaste bedrag kan worden beschouwd als een investering in de sector. Het bedrag wordt vooraf aan alle instellingen toegekend en is gebaseerd op het aantal in het peiljaar ingeschreven studenten. Het variabele deel kan gezien worden als een vorm van prestatiebeloning. Instellingen krijgen een bijdrage uit deze prestatiebox op basis van geleverde prestaties: hoe beter de prestatie, hoe hoger de bijdrage.

In een bestuursakkoord met de sectorraad (MBO Raad) zijn de doelen vastgesteld waarvoor de prestatiebox mag worden ingezet. Deze afspraken betreffen onder andere het terugdringen van het voortijdig schoolverlaten in het mbo en de tevredenheid van bedrijfsleven met de opleidingen en afgestudeerden. Het is de bedoeling dat de tevredenheid van het bedrijfsleven een van de grondslagen wordt voor de bekostiging van het mbo.

In totaal gaat het in het mbo om vier doelen, waarbij het eerste doel het meest concreet en uitgewerkt is:

- reductie van het aantal voortijdig schoolverlaters;
- tevredenheid van het bedrijfsleven;
- intensivering op het gebied van taal en rekenen;
- professionalisering van schoolteams.

De doelen worden gefaseerd ingevoerd. Dat traject wordt mede bepaald door het afbouwen van bestaande subsidies die voortaan in de prestatiebox zullen worden ondergebracht. In november 2011 is het bestuursakkoord over de professionaliseringsdoelen getekend door de staatssecretaris en de MBO Raad. Het akkoord bevat onder meer afspraken over de verhoging van de kwaliteit van het onderwijspersoneel en de bekwaamheid van het management en over de verbetering van het personeelsbeleid. Het mbo zal zich inzetten om het aantal masteropgeleide leraren substantieel te verhogen. De bekostiging van de professionaliseringsdoelstelling vindt plaats vanaf 2013. In 2012 wordt gestart met de verdere aanpak van het voortijdig schoolverlaten. In 2014 worden

de eisen ten aanzien van taal en rekenen ingevoerd. Wat betreft de tevredenheid van het bedrijfsleven (over de opleiding en over de afgestudeerden) verkeert men nog in de pilotfase.

De vier doelen mogen separaat worden beschouwd: men mag kiezen voor vier of voor minder doelen en er zit een schot tussen de middelen voor het eerste doel en de middelen voor de overige doelen. De prestatiebox wordt wel als één geheel uitgekeerd aan de instellingen, met zoals eerder gezegd een vaste en een variabele component. De afrekening vindt plaats op instellingsniveau. Een instelling heeft bijvoorbeeld recht op prestatiesubsidie als een of meer van de voortijdigschoolverlatennormen zijn gehaald dan wel als een substantiële daling in het uitvalpercentage wordt gerealiseerd. De school doet verslag van de inspanningen in het jaarverslag volgens een uniform format.

De maatregel wordt niet geïsoleerd genomen. Andere beleidsinterventies in het mbo betreffen onder andere: verkorting van de studieduur van mbo-4 naar drie jaar, herstructurering van mbo-1, afschaffing van de drempelloze instroom in mbo-2, verhoging van het aantal contacturen, het wettelijk vastleggen van onderwijs- en examenregeling, betere examens, centrale examinering in taal en rekenen (later voor mbo-4 ook Engels) en standaardisering van beroepsgerichte examens.

5.3 Doel

De prestatiebox is bedoeld om via een vereenvoudigde extra bekostiging landelijke doelen te realiseren wat betreft onderwijskwaliteit en onderwijsprestaties en om instellingen binnen de gestelde beleidskaders en -doelen voldoende ruimte te bieden daarbij eigen keuzes te maken en accenten te leggen. De uiteindelijke doelen van de maatregel zijn de verbetering van de onderwijskwaliteit en -prestaties, de daling van de ongediplomeerde uitstroom en de verhoging van de tevredenheid van het bedrijfsleven. Instellingen krijgen de middelen uit de prestatiebox boven op de lumpsumuitkering om zich in te zetten voor prestatiedoelen die in een bestuursakkoord op sectorniveau zijn of nog zullen worden vastgelegd. Het vaste deel van de prestatiebox betreft een investering in de sector waarvan het onderwijs beter moet worden. Het variabele deel is een vorm van prestatiebeloning. De prestatiebox is tevens bedoeld om een aantal bestaande subsidieregelingen te bundelen en te vereenvoudigen.

5.4 Veronderstellingen over gedragsreacties en effecten

Het beleid bevindt zich nog in een beginfase: delen van de maatregel zijn al uitgewerkt, andere nog niet. De invoering vindt gefaseerd plaats, hetgeen voor instellingen organisatorisch mogelijk eenvoudiger is. Het beleid is nog weinig concreet, zodat het niet goed mogelijk is om een uitspraak te doen over alle veronderstellingen over gedragsreacties en effecten.

Wel is een algemene veronderstelling dat de mbo-instellingen een kwaliteitsverbetering behoeven en dat een extra investering in plus een extra uitdaging aan de instellingen noodzakelijk zijn om tot het gewenste resultaat te komen. Tezamen moeten het

bestuursakkoord en de extra middelen het commitment van de sector en de individuele instellingen voor de landelijke beleidsdoelstellingen bewerkstelligen. Bovendien wil men naar een (deels) andere financieringsgrondslag, waarin meer aandacht is voor de tevredenheid van de stakeholders – in dit geval het bedrijfsleven.

Een andere veronderstelling is dat door een vaste beloningscomponent de kwaliteit bij alle instellingen zal verbeteren en dat door een variabel deel op te nemen instellingen (extra) geprikkeld raken om zich in te zetten voor de landelijke beleidsdoelen. Anders dan in het po en het vo kunnen instellingen in het mbo echt worden afgerekend op hun prestaties. In positieve zin kunnen ze hun budget verhogen door zich extra voor de gestelde doelen in te spannen.

5.5 Evaluatie

De centrale evaluatievragen zijn:

Wat betreft de opbrengstdoelen:

- 1 Treden na invoering van de prestatiebox veranderingen op in de prestaties en onderwijskwaliteit, het voortijdig schoolverlaten in het mbo en de arbeidsmarktpositie van afgestudeerde mbo-studenten?
- 2 Zijn de veranderingen overtuigend toe te schrijven aan de maatregel?

Wat betreft de instrumentele doelen:

- 3 Treden er veranderingen op in de mate van professionaliteit van management en docenten, de betrokkenheid van ouders, kwaliteit van het HRM-beleid en prestatiebeloning?
- 4 Is dit te danken aan de maatregel?

Wat betreft de sturingswijze:

- 5 Is deze vorm van financiering geschikt om de gestelde doelen te bereiken?

Effectevaluatie

Doordat het beleid op onderdelen nog weinig concreet is, is een goede effectmeting uitgesloten. Duidelijk is al wel dat het beleid voor alle instellingen tegelijk van kracht wordt. De doelstellingen worden gefaseerd, maar voor alle instellingen tegelijk ingevoerd. De effectiviteitsvraag lijkt vooralsnog niet te beantwoorden. Wel is het relevant om de ontwikkelingen in het mbo te blijven monitoren, en daarin wordt reeds voorzien.

Verklarende evaluatie

Andermaal moeten we een kanttekening maken bij de term verklarende evaluatie, die hier niet echt van toepassing is. Het beleid lijkt op dit moment vooral gebaat bij een studie naar de implementatie van het beleid. Hoe landt het beleid in de instellingen, welke waarde hechten bestuurders en directies aan het vaste, respectievelijk het variabele deel van de bekostiging en tot welk gedrag leidt dit op instellingsniveau? Om het voortijdig schoolverlaten terug te dringen bestond al een subsidiemaatregel. Gaat er dan nog wel

een prikkel uit van de prestatiebox? Welke concrete acties ondernemen instellingen om de professionalisering van het management en de docenten ter hand te nemen? Welke motieven voeren ze aan voor de gemaakte keuzes?

Daarnaast is het zinvol een onderzoek te starten naar het instrument prestatiebox. Is dit volgens de betrokkenen een goed sturingsinstrument om doelen te realiseren in het mbo? Voor welke doelen is de prestatiebox wel geëigend, voor welke doelen niet of veel minder? Is de snit (vast-variabel) van het instrument geschikt? Wat zijn de lessen voor de toekomst? Zo'n onderzoek zou via focusgroepen, Delphi-onderzoek dan wel een andere vorm van expertraadpleging kunnen plaatsvinden. Die variant van onderzoek lijkt ook het meest geschikt om de zin (volgens deskundigen en betrokkenen) te onderzoeken van een meer op tevredenheid van het bedrijfsleven gebaseerde bekostiging.

6 Introductie van entreeopleidingen in het mbo en afschaffing van de drempelloze instroom in mbo-2

6.1 Achtergrond en aanleiding

Mbo-1 heeft een zeer diverse instroom van studenten, onder wie op dit moment ook studenten die er qua intelligentie/leervermogen eigenlijk niet thuishoren. Het zijn vaak studenten met ‘gestapelde’ problemen. De aanwezigheid van veel studenten met problemen heeft geleid tot een vorm van onderwijs die vermengd is geraakt met opvang en hulpverlening. Sommige jongeren blijven te lang op school hangen zonder zich in te spannen voor hun studie. Dat maakt het werk van de docenten er niet eenvoudiger of aantrekkelijker op, en is bovendien storend voor de studenten die wel gemotiveerd zijn om te studeren.

De problemen beperken zich niet tot het mbo-1. Het gehele mbo kampt met veel voortijdige schooluitval. Ook kwam het mbo het afgelopen jaar soms negatief in het nieuws door klachten van studenten over te weinig lessen en contacturen.

Het niet-stellen van vooropleidingseisen voor toelating tot een mbo-2-opleiding komt de kwaliteit van het mbo-2-onderwijs niet ten goede. De drempelloze instroom belemmert een goede aansluiting bij het overige mbo-onderwijs.

Het kabinet heeft maatregelen getroffen om de kwaliteit van het mbo-onderwijs te verbeteren. Twee van die maatregelen zijn nauw met elkaar verweven en staan hier centraal. Het gaat om het vervangen van mbo-1 door de zogenoemde entreeopleiding en het afschaffen van de drempelloze instroom in mbo-2.

6.2 Inhoud

In 2013/2014 worden de maatregelen naar verwachting van kracht. Vanaf dat jaar worden de mbo-1 en aka-opleidingen ‘entreeopleiding’. De entreeopleiding is bedoeld voor studenten die niet voldoen aan de eisen voor mbo-2, -3 of -4 en biedt entree tot een mbo-2-opleiding of tot de arbeidsmarkt. Afstromen vanuit mbo-2, -3 of -4 naar de entreeopleiding is niet mogelijk. Voor de mbo-2-opleidingen gelden wettelijke vooropleidingseisen, de drempelloze instroom wordt afgeschaft.

Voordat een student wordt toegelaten tot de entreeopleiding vindt een intake plaats, waarin kan worden besloten om de student te testen en hem bij positieve uitkomst alsnog toe te laten tot een mbo-2-, -3- of -4-opleiding.

Iedere student in de entreeopleiding krijgt na vier maanden een studieadvies. Als de student niet de gewenste studieresultaten heeft geboekt, kan dat een bindend negatief advies zijn. De student moet de opleiding dan beëindigen. Het studieadvies is omgeven door een zorgvuldige procedure. Voorafgaand aan een negatief advies wordt een waarschuwing gegeven. De student kan tegen het besluit in beroep gaan. Studenten die jonger zijn dan achttien jaar krijgen bij een negatief studieadvies de raad om het in een andere (entree-)opleiding of bij een andere instelling te proberen. Zo wordt geprobeerd te voorkomen dat zij voortijdig schoolverlaters worden. De instelling kan via een

onderwijsovereenkomst – een contract tussen de instelling en student – verdere afspraken regelen over prestaties en bijvoorbeeld aanwezigheidseisen.

Studenten in de entreeopleiding die met meervoudige problematiek kampen, kunnen naar een lokale of regionale plusvoorziening worden verwezen, waar zij – naast hun mbo-onderwijs – worden geholpen met hun problemen. De plusvoorziening is bedoeld voor studenten van 12 tot 23 jaar.

In de entreeopleiding komt meer geld per student beschikbaar dan nu. De bekostiging wordt separaat van mbo-2, -3 en -4 berekend. Het bedrag voor de entreeopleiding maakt deel uit van de lumpsum, maar de hoogte van het bedrag wordt apart vermeld in de bekostigingsbrief aan de instellingen. Een student in de entreeopleiding wordt maximaal voor twee jaar bekostigd, dit is tevens de maximale inschrijvingsduur, na twee jaar kan een student dus geweigerd worden. De bekostiging is aflopend (cascade), wat wil zeggen dat de bekostiging voor het eerste jaar hoger is dan die voor het tweede jaar. Het onderwijsprogramma in de entreeopleiding duurt één jaar en telt minimaal duizend uur, waarvan minimaal zeshonderd uur ‘begeleide onderwijstijd’.

De maatregel past binnen het beleid van dit kabinet voor verhoging van de kwaliteit, vereenvoudiging van het bve-stelsel en verbetering van de besturing en bedrijfsvoering (Focus op Vakmanschap 2011-2015). De maatregel vindt niet in een isolement plaats. Het beleid is erop gericht om alle opleidingen in het mbo te verkorten en te intensiveren, de schoolse component te versterken, meer eisen te stellen op het gebied van taal en rekenen (tot niveau 2F resp. 3F van de referentiekaders voor taal en rekenen) en de betrokkenheid van ouders bij het onderwijs van hun zoon of dochter te vergroten.

6.3 Doel

Door de invoering van entreeopleidingen krijgen studenten zonder diploma de mogelijkheid om een opleiding op maat te volgen die toegang geeft tot een mbo-2-opleiding of de arbeidsmarkt. De maatregel om entreeopleidingen in te voeren in het mbo kan niet los worden gezien van het afschaffen van de drempelloze instroom in mbo-2. Beide maatregelen zijn bedoeld om de kwaliteit van het beroepsonderwijs een positieve impuls te geven. Dat gebeurt door eisen te stellen aan de vooropleiding voor mbo-2 en door opvang en hulpverlening te scheiden van het onderwijs. De instellingen krijgen via een zogenoemd bindend studieadvies de mogelijkheid zich eerder te ontdoen van niet-gemotiveerde en niet-presterende studenten. De maatregel moet leiden tot meer gediplomeerde uitstroom en minder schooluitval in mbo-1 en 2 en tot het vergroten van de aantrekkelijkheid van het mbo.

6.4 Veronderstellingen over gedragsreacties en effecten

De veronderstelling is dat het huidige aannamebeleid voor mbo-1 en de drempelloze instroom in mbo-2 de kwaliteit van het mbo-onderwijs aantasten. De kwaliteit van het mbo-onderwijs staat eveneens onder druk doordat studenten met meervoudige problemen niet adequaat worden geholpen en doordat de instellingen daarnaast niet-gemotiveerde studenten te lang moeten tolereren.

De verwachting is dat mede als gevolg van de maatregel de kwaliteit van het mbo toeneemt, de prestaties en het rendement in het mbo omhooggaan en het voortijdig schoolverlaten afneemt. Door parallel beleid met zwaardere eisen op het gebied van taal en rekenen kunnen de rendementseffecten echter getemperd worden.

Effecten op instellingsniveau

De verwachting is dat door een verbeterde intake in de entreeopleiding en door wettelijke vooropleidingseisen aan mbo-2 studenten beter worden voorgesorteerd. De samenstelling van de groepen in de entreeopleiding, maar ook in mbo-2 zal als gevolg van de maatregel homogener zijn.

De verwachting is dat door tijdige verwijzing naar een plusvoorziening het onderwijs niet of minder gehinderd wordt door studenten met meervoudige problematiek. Door de mogelijkheid om studenten een negatief bindend advies te geven, heeft de instelling bovendien minder of geen hinder van niet-presterende studenten.

Een andere verwachting betreft de cascade: deze financieringswijze moet instellingen prikkelen tot meer rendement.

Effecten op studentenniveau

Studenten in het huidige mbo-1 krijgen niet altijd het juiste onderwijs. De veronderstelling is dat studenten dankzij de verbeterde intake – met de mogelijkheid om het niveau van de student te testen en de student eventueel alsnog drempelloos te laten instromen in andere mbo-niveaus – meer onderwijs op maat krijgen en vaker op het niveau zullen studeren waarvoor ze de capaciteiten hebben.

De maatregel werkt naar verwachting ook door op het gedrag van studenten. Een positief studieadvies na vier maanden zal aanmoedigend werken en de motivatie van studenten versterken. Het is mogelijk dat studenten na dit gesprek hun keuze voor een opleiding definitief maken en dat er na die vier maanden minder vaak gewijzigd zal worden.

De goedwillende studenten zullen bovendien niet langer lijden onder de ongemotiveerde, niet-presterende studenten, aangezien deze tijdig een bindend negatief studieadvies kunnen krijgen. Het studieplezier van de overgebleven studenten zal als gevolg daarvan toenemen.

De sanctiemogelijkheid van een negatief bindend studieadvies zet studenten mogelijk aan zich meer in te spannen voor hun studie. Door studenten met problemen door te verwijzen naar de plusvoorziening zullen deze studenten eerder en beter worden geholpen, hetgeen hun (onderwijs)loopbaan ten goede komt.

Effecten op docentniveau

Een andere veronderstelling betreft de docenten. De meer homogene groepen, het scheiden van onderwijs en opvang/hulpverlening en het tijdig kunnen weren van niet-gemotiveerde, niet-presterende studenten maken het lesgeven voor docenten naar verwachting eenvoudiger en aantrekkelijker.

Effecten op de relatie tussen instelling en gemeente

De samenwerking tussen gemeente en instellingen op het vlak van opvang en hulp van jongeren met meervoudige problematiek wordt mogelijk beter. Een andere verwachting betreft de transparantie in de bekostiging van de entreeopleiding. Deze maakt het eenvoudiger voor instellingen om aan de gemeente zichtbaar te maken hoeveel ruimte er binnen de entreeopleiding is.

Onbedoelde effecten en beleidsafwijkingen

Mogelijk doen zich ook onbedoelde, negatieve effecten voor. Een van die effecten zou het doorschuiven van de hete aardappel kunnen zijn: instellingen die wellicht te snel besluiten om studenten na vier maanden te verwijderen met de raadgeving het eens elders te proberen. De kans dat een dergelijk effect optreedt, wordt vergroot door beleid dat erop gericht is om alle jongeren eerst een opleiding te laten volgen. Dit beleid leidt wellicht tot een (te) grote instroom van weinig gemotiveerde jongeren in de entreeopleiding. Een ander onbedoeld effect is een mogelijk vergaande juridisering rond de bindendadviesprocedure.

6.5 Evaluatie

Hoe kan worden vastgesteld of de beoogde effecten door de combinatie van maatregelen worden bereikt? De evaluatie zal gericht moeten zijn op zowel het mbo-1 als het mbo-2 en in kaart moeten brengen welke ontwikkelingen zich hier – op verschillende niveaus – als gevolg van deze interventie voordoen. Het betreft veranderingen in de omvang van mbo-1 en mbo-2, in de samenstelling en in de prestaties. Daarnaast is het relevant te kijken tot welke gedragsveranderingen de maatregel leidt bij instellingen, docenten, studenten en eventueel ook gemeenten.

Effectevaluatie

Noch het random gecontroleerde experiment, noch het natuurlijk experiment is bij deze maatregel van toepassing. De maatregel wordt integraal ingevoerd op alle instellingen, voor alle opleidingen. Er is dus geen onderscheid tussen een experimentele en een controlegroep. De maatregel wordt niet vooraf in pilots of experimenten getest. Het is niet goed mogelijk om vergelijkbare groepen te vinden voor een experimentele opzet. Bovendien verandert er veel tegelijkertijd, waardoor een common-trendsassumptie evenmin houdbaar lijkt.

De maatregel laat zich waarschijnlijk het best evalueren door een cohortvergelijking voor en na de invoering van de entreeopleiding en de wettelijke toelatingseisen voor mbo-2.

De maatregel zal tot een andere verdeling van studenten over de beide niveaus leiden. Een eerste vraag in deze effectevaluatie is hoe de omvang van mbo-1 en mbo-2 zich ontwikkelt als gevolg van de maatregel. Een tweede vraag luidt of de samenstelling van entree- en mbo-2-opleidingen meer homogeen is dan voorheen. Een derde cluster van vragen richt zich op de schoolprestaties en de arbeidsmarktpositie van vergelijkbare studenten onder het oude en het nieuwe regiem. Belangrijke indicatoren in de

cohortvergelijking zijn: instroom-, doorstroom- en gekwalificeerde uitstroomgegevens (diplomering), schooluitval en de arbeidsmarktpositie van uitgestroomde studenten.

Het is de vraag hoe vergelijkbare studenten in het mbo presteren voor en na invoering van de beleidsmaatregel. Bij de cohortvergelijking moet zo goed mogelijk rekening worden gehouden met relevante factoren en omstandigheden om een zo zuiver mogelijke vergelijking te maken tussen de situatie voor en de situatie na de invoering van de maatregel.

Een belangrijke voorwaarde voor zo'n vergelijking is de beschikbaarheid van voldoende gegevens. Deze kunnen worden gebaseerd op bestaande BRON-gegevens. Een voorwaarde is dat de studenten vergelijkbaar zijn naar achtergrondkenmerken (bv. etniciteit) en voorgeschiedenis (niveau, diploma). Bij DUO moet worden nagevraagd welke mate van verfijning op dit punt mogelijk is. Aanvullend kunnen de gegevens worden verrijkt met gegevens uit het SSB (CBS).

De analyse van meerdere cohorten voor en na invoering van de maatregel kan dan iets zeggen over het effect van de maatregel in termen van studierendement (voortgang, doorstroom, diplomering en uitval). Met behulp van de ROA-data kan een vergelijking worden gemaakt van de arbeidsmarktpositie van jongeren onder het oude en onder het nieuwe regiem.

Het is evenwel belangrijk om voor ogen te houden dat waargenomen veranderingen niet uitsluitend en zonder meer kunnen worden toegeschreven aan de maatregel. Boven genoemde vergelijkingen geven niet meer dan een grove schatting van het effect van het beleid. De maatregel is immers slechts een van de beleidsinterventies waarmee het mbo te maken krijgt, en bovendien moet rekening worden gehouden met cohortverschillen.

Verklarende evaluatie

De verklarende evaluatie is bedoeld om de mechanismen te achterhalen waardoor de maatregel wel, niet of niet in alle opzichten effect heeft. Centrale vraag voor de verklarende evaluatie is dan ook welke veranderingen als gevolg van de maatregel op instellingsniveau optreden, bij docenten en bij studenten, en welke andere factoren de implementatie en uitvoering van de maatregel in de praktijk beïnvloeden in positieve dan wel negatieve zin.

Behalve naar processen op instellingsniveau is het relevant te kijken welke veranderingen zich lokaal voordoen in de samenwerking tussen gemeente en instellingen om jongeren (al dan niet met problemen) aan een opleiding te helpen.

Er zijn verschillende mogelijkheden om meer zicht te krijgen op de wijze waarop het beleid van invloed is op het gedrag van instellingen, docenten, studenten en gemeenten. Daarbij kunnen (kwantitatief) bestaande monitoren, ex ante en ex post peilingen en (kwalitatief) gerichte interviews dienstdoen.

Op instellingsniveau is het interessant te onderzoeken hoe men de nieuwe intake organiseert en deze waardeert ten opzichte van de oude werkwijze. Zijn de instellingen in staat tot een adequate voorsortering van studenten? Klopt de veronderstelling dat groepen nu homogener (gelijk niveau en weinig persoonlijke problemen) zijn? Hoe verhoudt

de nieuwe maatregel zich tot in het verleden gehanteerde rendementsbevorderende maatregelen?

Bij voorkeur door middel van een ex ante en ex post vergelijking kan tevens de tevredenheid van docenten worden gepeild. Maakt de voorsortering van studenten het lesgeven inderdaad eenvoudiger en prettiger?

Voor het globaal peilen van studententevredenheid kan de JOB-monitor worden ingezet.

7 Centrale examinering van taal en rekenen in het mbo

7.1 Achtergrond en aanleiding

Op dit moment kent het mbo geen centrale examinering en zijn de instellingen zelf verantwoordelijk voor de examens. Dat heeft tot wisselende kwaliteit van examinering geleid, en bij sommige opleidingen tot ondermaatse examens. Op 1 februari 2012 constateerde de Inspectie van het Onderwijs dat er 19 zeer zwakke mbo-opleidingen zijn, plus nog 83 mbo-opleidingen waarvan de examinering onder de maat is.

Eerder al heeft het kabinet maatregelen aangekondigd om de kwaliteit van de eind-examens in het mbo te verbeteren. Als opleidingen op het punt van examenkwaliteit in gebreke blijven, zal dat openbaar worden gemaakt en worden bestraft met bestuurlijke sancties. Als uiterste maatregel kan de minister besluiten tot intrekking van de licentie van zulke opleidingen (Focus op Vakmanschap). De introductie van centrale examens is onderdeel van de verzwaring van de kwaliteitseisen en moet leiden tot een structurele kwaliteitsverbetering.

In het onderwijsbeleid is (in alle sectoren) de laatste jaren meer nadruk komen te liggen op taal en rekenen, en dat geldt ook voor het mbo. In 2010 zijn de referentieniveaus voor taal en rekenen ingevoerd. De komende jaren zullen in het mbo centrale examens voor taal en rekenen verplicht worden. Centrale examinering in taal en rekenen kan als het sluitstuk van de introductie van de referentieniveaus worden gezien.

7.2 Inhoud

De centrale examinering in het mbo heeft een lange invoeringstijd. De verplichting gaat in 2013-2014 gelden voor mbo-4, waarvoor het referentieniveau 3F geldt, dat gelijkstaat aan havoniveau. In 2014-2015 wordt centrale examinering verplicht voor mbo-2 en mbo-3, waarvoor het referentieniveau 2F geldt. In 2017-2018 wordt voor mbo-4 ook een centraal examen Engels verplicht. Het College voor Examens (CvE) krijgt de regie over de centrale examinering.

Het is niet zo dat de exameneisen tussentijds voor studenten worden verzwaard. De kwalificatie-eisen bij opleidingen en beginnende studenten zijn al bekend sinds 2010, toen de referentieniveaus werden geïntroduceerd. Studenten uit dat jaar krijgen vanaf 2013-2014 met de centrale examinering te maken. Studenten die voor 1 augustus 2010 aan hun studie begonnen, mogen afstuderen volgens de regels van het oude regiem.

De examinering vindt digitaal plaats. De digitale centraal ontwikkelde examens worden op de scholen afgenomen in drie periodes van twee weken. Binnen die aangewezen periode kan de school de digitale toetsen afnemen op een manier die past in het rooster. Er is ook een mogelijkheid van een papieren toets, deze wordt op één centraal aangewezen tijdstip binnen iedere periode afgenomen. Elke deelnemer mag binnen het aanbod van de school zelf bepalen wanneer het centraal ontwikkelde examen afgelegd wordt. De deelnemer kan maximaal één keer per periode het centraal ontwikkelde examen afleggen. Een school die kiest voor digitale afname van de centraal ontwikkelde

examens heeft dus aanmerkelijk meer speelruimte dan een school die kiest voor papieren centraal ontwikkelde examens. Uitgangspunt is dat er van ieder centraal ontwikkeld examen taal, respectievelijk rekenen, meerdere versies komen. Dit is bedoeld om te waarborgen dat er niet gefraudeerd wordt door instellingen en studenten. Er is getest, zodat de varianten alle van gelijke zwaarte zijn. Een deelnemer in de sector techniek legt dus een ander centraal ontwikkeld examen af dan een deelnemer in de sector welzijn. In de versies van de centraal ontwikkelde examens worden wel dezelfde vaardigheden getoetst, maar de versies verschillen op de contexten waarin de opgaven worden aangeboden.

Bepaalde examenopgaven komen voor in zowel het centraal eindexamen havo als in het examen mbo-4, zodat het niveau van beide opleidingen met elkaar kan worden vergeleken.

Pilots

Het beleid verkeert momenteel in een pilotfase. In 2011-2012 vond een pilot plaats voor niveau 3F (mbo-4) en een pre-pilot voor niveau 2F (mbo-2 en -3). De pilots hebben verschillende doelen. Ze beogen de instellingen te laten wennen aan de centrale examinering, de docenten bekend te maken met de inhoud van de examens (en wat ze dus aan taal en rekenen moeten onderwijzen), zicht te krijgen op het prestatieniveau van de studenten en de normering te bepalen. De deelname aan de pilots is vrijwillig: alle instellingen mogen meedoen, maar niet met alle studenten. Het enthousiasme voor deelname aan de (gratis) pilots is groot.

De pilots werden voor het eerst uitgevoerd in februari 2011. Toen nam de helft van de instellingen voor het mbo-4 met een deel van de studenten aan de pilots deel.

In het tweede jaar van de pilots, 2012, wordt twee keer per jaar (januari en mei/juni) een centraal eindexamen afgenomen. Waarschijnlijk zullen alle instellingen dan meedoen, maar ook weer niet met alle studenten. Het laatste jaar van de pilot, 2013, wordt gezien als een generale repetitie: dan zullen alle instellingen, met alle studenten, meedoen. Mocht blijken dat de consequenties voor de eindexamenresultaten dramatisch zullen zijn, dan kan tot uitstel worden besloten (en dus tot verlenging van de pilotfase).

Voorafgaand aan de pilots zijn er aan het einde van de opleiding metingen verricht om het referentieniveau te toetsen. Die metingen hebben plaatsgevonden in 2009, 2010 en 2011. Tussentijds is de manier van toetsen aangepast, waardoor de inhoud en de normering niet helemaal vergelijkbaar zijn. Hierdoor kan ook niet echt worden gesproken van een nulmeting in 2009. De pilotexamens zijn door het Cito evenwel terug te vertalen naar de eerdere metingen. Door selecties uit de deelnemers aan de pilots wordt een representatief beeld gegeven van de sector. Het CVE heeft de regie over alle peilingen.

7.3 Doel

Passend bij een grotere nadruk op taal en rekenen in het gehele onderwijs, met referentieniveaus die een doorgaande leerlijn moeten waarborgen, wordt in het mbo een centrale examinering voor deze vakken geïntroduceerd. Hoofddoel van de centrale

examinering in taal en rekenen is het verbeteren van de kwaliteit van het taal- en rekenonderwijs in het mbo en daarmee van de prestaties. Ook helpt zij de waarde of kwaliteit van diploma's in het mbo te garanderen. De maatregel beoogt eveneens de doorstroom van mbo-4 naar het hbo beter te laten verlopen. Centrale examinering zorgt er verder voor dat het vergelijken van prestaties bij verschillende instellingen mogelijk wordt.

In de toekomst moeten studenten in het mbo-2, -3, en -4 centraal examen doen in taal- en rekenvaardigheden. De verplichting verloopt gefaseerd voor de verschillende niveaus in het mbo. Voor deze maatregel is 50 miljoen euro per jaar ter beschikking gesteld.

7.4 Veronderstellingen over gedragsreacties en effecten

De beleidstheorie veronderstelt dat centrale examinering een sterke prikkel is voor instellingen om meer aandacht te schenken aan taal en rekenen en tot betere prestaties te komen. Het beleid mag succesvol worden genoemd als de prestaties verbeteren. Nog beter is het als na invoering evenveel mbo-studenten slagen als voor invoering van centrale examinering. Naar het zich laat aanzien, zal dat niet het geval zijn: veel studenten in de pilots hebben grote moeite met de taal- en rekentoetsen. De instellingen vrezen de uitslag van de echte examinering. Mogelijk lag de lat in de pilots te hoog en moet de norm naar beneden worden bijgesteld.

De veronderstelling is dat een diploma met centrale examinering meer waard is dan een diploma zonder taal- en rekenexaminering en/of met lokaal ontwikkelde examens. Een andere veronderstelling is dat de doorstroming naar het hbo voor de mbo-4-studenten makkelijk zal verlopen (en dus tot minder uitval zal leiden) bij een goede beheersing van taal en rekenen (en later: Engels).

Effecten op instellingsniveau

De veronderstelling is dat er veel meer aandacht gaat worden besteed aan taal en rekenen in alle opleidingen binnen een instelling. Die toegenomen aandacht zou zich moeten vertalen in betere prestaties in die vakken en in gediplomeerde uitstroom. Instellingen kunnen het zich niet permitteren om met grote aantallen voor het examen gezakte studenten in de publiciteit of ranglijsten te komen. Zulke publiciteit schaadt het imago van de opleiding/instelling. Instellingen of opleidingen daarbinnen zullen daarom zo goed mogelijk proberen in te spelen op de strengere exameneisen, bijvoorbeeld door hun docenten te laten bijscholen op beide vakken. Zo lanceerde de gemeente Rotterdam recent een plan om alle docenten in het vo en het mbo een taaltoets te laten afleggen om eventuele deficiënten in hun kennis op te sporen.

Effecten op studentenniveau

Een van de veronderstellingen is dat een mbo-student die centraal examen in taal en rekenen heeft gedaan vaardiger is dan een mbo-student die dat niet heeft gedaan. Dat effect wordt niet alleen gecreëerd door de examinering op zich, maar naar verwachting ook door een toenemend besef bij studenten dat dit belangrijke (doorstroom/eind-examen) vakken zijn, waardoor ze zich hiervoor meer zullen inspannen.

De veronderstelling is dat mbo-studenten die slagen voor het examen beter zullen doormaken naar het hbo, en dat hun start op de arbeidsmarkt zich gunstig onderscheidt van de afgestudeerden oude stijl en uiteraard ook van degenen die niet in staat waren de examens met goed gevolg af te leggen.

Als gevolg van de maatregel zullen verschuivingen optreden in de samenstelling van de studentenpopulatie naar type opleiding en niveau. Studenten die moeite hebben met taal en rekenen zullen eerder voor een lager niveau en type opleiding kiezen.

Effecten op docentniveau

Docenten zullen zich gezien het toenemend belang dat wordt gehecht aan betere taal- en rekenprestaties bij studenten meer gaan inspannen om hun studenten op het gewenste niveau te brengen. Het is vooralsnog de vraag of dit met name de kwaliteit van de docenten Nederlands en wiskunde/rekenen betreft of dat deze inzet integraal, van het hele team van docenten, verlangd wordt. Mogelijk verschilt dat per instelling of opleiding.

Onbedoelde effecten of beleidsafwijkingen

Een van de doelen van het kabinet is de aantrekkelijkheid van de beroepsgerichte route in het onderwijs te vergroten, of anders gezegd: de dominantie van de algemeen vormende route te doorbreken (Focus op vakmanschap). Het is de vraag of het zwaardere examineren in taal en rekenen daaraan bijdraagt, of voor welke studenten dat wel en voor welke studenten dat niet het geval is. Te veel nadruk op cognitieve schrijft mogelijk bepaalde studenten af.

Waarschijnlijk zijn niet alle instellingen/opleidingen/docenten voorstander van meer nadruk op cognitieve/algemene vakken. Daarbij is er wellicht nog verschil in houding ten aanzien van de bol-opleidingen en de bbl-opleidingen. Als bij de docenten (in welke route – bol of bbl – dan ook) geen eigenaarschap optreedt van dit beleidsdoel, dan zal het taal- en rekenonderwijs zich niet verder ontwikkelen en zal de kwaliteit van het onderwijs op deze terreinen niet toenemen.

Een andere kwestie betreft de vaardigheden van docenten zelf. Bij lang niet alle docenten zullen de Nederlandse taal- en rekenvaardigheden op orde zijn. Zulke docenten kunnen uiteraard moeilijk hun studenten goed voorbereiden. Die docenten laten bijscholen op beide terreinen lijkt gewenst, maar dat is aan tijd, kosten en wellicht ook andere grenzen gebonden.

Daarnaast zou het zo kunnen zijn dat goede vakdocenten/praktijkdocenten verloren gaan voor het mbo als van hen wordt verwacht dat ze studenten taal- en rekenvaardigheden bijbrengen.

Een andere mogelijk ongewenste reactie als gevolg van de maatregel zou kunnen zijn dat studenten 'afhakken'. Sommige studenten in het mbo die wel gemotiveerd zijn voor hun opleiding zullen de nieuwe eisen op het gebied van taal en rekenen als een verzwarende van de studie zien, hetgeen hun motivatie negatief beïnvloedt. Afstroom naar lagere niveaus of voortijdig schoolverlaten zijn bijbehorende risico's. Naar verwachting

zullen deze effecten zich sterker voordoen bij studenten in de bbl-opleidingen dan bij studenten in de bol-opleidingen.

Een variant van dit verschijnsel is als studenten niet langer de gok durven wagen naar een hoger niveau.

7.5 Evaluatie

Effectevaluatie

Bij deze maatregel zijn geen experimentele groepen en controlegroepen te onderscheiden. Wordt de maatregel van kracht, dan gebeurt dat integraal en geldt deze voor alle mbo-scholen. Er zijn weliswaar pilots gestart, maar deze zijn niet opgezet als gerandomiseerd experiment. Bovendien is een pilot niet gelijk aan de feitelijke maatregel. De hier gestarte en nog te starten pilots zijn met name bedoeld om instellingen te laten wennen aan centrale examinering en om de normering te kunnen bepalen. Wel kunnen de pilots worden benut als nulmeting om de prestaties van vergelijkbare studenten voor invoering van de maatregel te vergelijken met die van na invoering van het feitelijk examen. We hebben het dan echter over het *monitoren* van ontwikkelingen en niet over een effectmeting. Voor zo'n vergelijking moeten door Cito de resultaten van eerdere metingen wellicht door weging eerst vergelijkbaar worden gemaakt. Na correctie/transformatie kan dan een schatting worden gemaakt van het effect van feitelijke examinering door de resultaten uit de oefenfase te vergelijken met die van de examinering, als eerste in 2013-2014 voor mbo-4. Er moet worden gecontroleerd voor achtergrondkenmerken. Andere zinvolle vergelijkingen op basis van monitorgegevens zijn die tussen leerlingen in havo-4 en studenten in mbo-4. Beide maken de toetsen immers op hetzelfde niveau.

Om toch te proberen het effect van de maatregel vast te stellen, bieden wellicht een paar varianten van de difference-in-differencebenadering mogelijkheden.

De maatregel wordt per niveau in het mbo voor alle studenten tegelijk ingevoerd, te beginnen met niveau 4. Toch is het voorstelbaar dat de maatregel voor sommige studenten in mbo-4 meer consequenties heeft dan voor andere. Studenten die geen of nauwelijks problemen met taal en rekenen hebben, zullen weinig problemen ondervinden van een centrale examinering. Voor deze studenten is het verwachte effect van het beleid gering of nihil, en we kunnen hen als controlegroep beschouwen. Daarentegen zijn er ook studenten voor wie de maatregel zeer relevant is: namelijk de studenten voor wie rekenen en taal veel problemen opleveren. Zij lopen een grotere kans onvoldoendes te halen bij centrale examinering en de maatregel heeft voor hen dus veel consequenties. Zij kunnen worden beschouwd als experimentele groep.

Eerder onderzoek liet zien dat rekenen en taal met name veel problemen opleveren voor mbo-studenten in de Randstad c.q. sterk verstedelijkt gebied. Een manier om vast te stellen of het beleid effect heeft, is een voor- en nameting verrichten bij de experimentele groep uit sterk verstedelijkte postcodegebieden en een controlegroep elders in het land. Een andere variant is de vergelijking tussen opleidingssectoren waarvoor de maatregel weinig of geen effect heeft (bewezen uit onderzoek dat studenten daar al goed zijn in

taal en rekenen) – de controlegroep – en sectoren waarvan bekend is dat studenten er moeite hebben met taal en rekenen – de experimentele groep). Weer een andere variant is wat betreft taalexaminering een onderscheid maken tussen meisjes (controlegroep) en jongens (experimentele groep), al dan niet in bepaalde opleidingen; of andersom wat betreft de rekenexamens. Het effect van de maatregel kan dan worden bepaald door het verschil in voor- en nameting in de controlegroep af te trekken van het verschil in voor- en nameting in de experimentele groep. Wel moeten we dan veronderstellen dat de trend bij de groepen hetzelfde zou zijn geweest als de beleidswijziging niet zou zijn doorgevoerd. Bij deze difference in differences is de aanname van gelijke trend cruciaal, maar de groepen waartussen vergeleken wordt, mogen wel van elkaar verschillen. Een andere optie is die van een trendvergelijking tussen havoleerlingen en mbo-4-studenten. In deze vergelijking geldt eveneens dat de pilots als voormeting worden beschouwd en de feitelijke examens als nameting. Door nu tussen de beide sectoren de trend te vergelijken kunnen we op het spoor komen van het effect van de maatregel. De veronderstelling is immers dat verder vergelijkbare studenten door beter onderwijs de examens beter zullen gaan maken. Als de verschillen tussen havo en mbo in de nameting kleiner zijn dan in de voormeting, dan zou dat een indicatie kunnen zijn van het effect van het beleid. Hierbij moet echter wel zeer kritisch worden gekeken naar de common-trendsassumptie.

Verklarende evaluatie

Een verklarende evaluatie is erop gericht de mechanismen te achterhalen waardoor een bepaalde interventie wel, geen of niet in alle omstandigheden effect heeft. Een verklarende evaluatie kan op verschillende momenten tijdens of na de beleidsimplementatie worden uitgevoerd. In dit geval kunnen we de pilots als onderdeel van een verklarende evaluatie beschouwen. Immers, deze zijn mede bedoeld om te kijken of instellingen erin slagen de examens te organiseren en docenten voor te bereiden op de inhoud van de examinering. Door ook studenten te bevragen naar hun voorbereiding op het proefexamen en hun ervaring daarmee krijgt men inzicht in het leerproces en de werking van examinering.

De pilots zullen bij een grondige evaluatie dan ook duidelijk kunnen maken in welke context de instellingen erin slagen de examinering goed te laten verlopen en wat de beste voorbereiding van docenten en studenten is.

Aangezien het pilots betreft, zou een verklarend onderzoek na invoering van de feitelijke examens ook interessant zijn. Een van de onderdelen van zo'n onderzoek is het analyseren van monitorgegevens om ontwikkelingen in de prestaties en de slagingspercentages van studenten per opleiding en instelling te volgen. Zo'n analyse kan verklaren welke studenten, op welk niveau en in welke opleiding en instelling meer of juist minder kans hebben om de examens in taal en rekenen met goed gevolg af te leggen.

In een selectie van instellingen en opleidingen zou vervolgens aan docenten en directies de vraag kunnen worden voorgelegd of men door de pilots goed was voorbereid, of dat de feitelijke examinering tot verrassingen heeft geleid. Een van de zinvolle vragen

die daarmee beantwoord kan worden, is of diagnostische toetsen, pilots en feitelijke examens door de betrokkenen verschillend worden voorbereid en beleefd.

Een ander onderzoek zou zich moeten richten op de verdere (school-)loopbaan van de mbo-ers. Is het inderdaad zo dat degenen die centrale examens hebben afgelegd voor taal en rekenen (en eventueel Engels) beter doorstromen in het hbo? Dit kan bijvoorbeeld worden bepaald middels cohortvergelijking. Vergelijkbare vragen hebben betrekking op de arbeidsmarktpositie van beide groepen studenten.

8 Standaardisering beroepsgerichte examens in het mbo

8.1 Achtergrond en aanleiding

Uit onderzoek van de Inspectie van het Onderwijs naar de kwaliteit van examens in het mbo is gebleken dat bij veel instellingen in het mbo de kwaliteit van de examens onvoldoende is. Volgens de laatste gegevens was bij 63% van de opleidingen het examen van voldoende kwaliteit, maar bij 37% dus nog niet. De inspectie ziet steekproefsgewijs toe op de examenkwaliteit. Dat is overigens pas enkele jaren het geval. Voorheen was dit toezicht een taak van het KCE, maar dat voldeed niet. De problemen met de examens zijn meervoudig: de dekking van de kwalificatiedossiers was onvoldoende, de moeilijkheidsgraad van de examens liet te wensen over en er waren problemen met de gebruikte toetsvormen en de gehanteerde beoordelingsmodellen. Voorheen waren de problemen groter bij de bekostigde instellingen, maar recent zijn het juist vaker de niet-bekostigde instellingen waar tekorten worden geconstateerd. De inspectie heeft geconstateerd dat bij veel opleidingen de kwaliteit van de examens en de kwaliteitszorg niet in orde zijn. Doorgaans voeren scholen na een beoordeling van de inspectie wel veranderingen door voor de opleidingen waar tekorten zijn geconstateerd, maar laten ze na de problemen integraal (voor alle opleidingen) aan te pakken.

Het idee is dat door standaardisering van de beroepsgerichte examens in het mbo de kwaliteit van de examens verbetert. Een kwestie die hiermee verbonden is, betreft het centraal inkopen van examens. Uit onderzoek van het RCEC (Sanders en Visser 2008) blijkt dat hieraan in vergelijking met het als instelling/opleiding zelf ontwikkelen van examens vele voordelen zijn:

- Het levert betere examens op.
- Het is doelmatiger (goedkoper).
- Het komt de transparantie ten goede.
- Het is eenvoudiger om het bedrijfsleven erbij te betrekken, en het bedrijfsleven heeft ook meer vertrouwen in zulke centraal ingekochte examens.
- Docenten houden meer tijd over voor het primaire proces: het onderwijs zelf.

Overigens bevonden zich volgens de inspectie ook onder de nu reeds centraal ingekochte examens examens van onvoldoende kwaliteit. Aanvankelijk werd gedacht aan de invoering van een keurmerk voor goede examens, maar dat wordt nu als een onbegaanbare weg beschouwd (te veel variatie, veel te veel opleidingen).

Inmiddels zijn verschillende sectoren al bezig met een vorm van standaardisering. In de zorg, met het oog op de beroepspraktijk, is men hiermee al vrij ver, maar in andere sectoren is het allemaal nog weinig concreet.

Er is een pilot geweest met slechts enkele instellingen die met de profielen aan de slag zijn gegaan. Het doel van de pilot was de instelling te laten wennen aan de profielen. De vraag was of de examenprofielen voor de scholen hanteerbaar waren. Wat betekenen die

voor onze examens en voor ons onderwijs, en welke aanpassingen zijn nodig? Nog voordat deze pilot was geëvalueerd, is het beleid landelijk ‘uitgerold’.

8.2 Inhoud

Het is de bedoeling dat in het mbo standaardisering van examens tot stand komt. De maatregel heeft betrekking op alle instellingen in het mbo: de roc's, de aoc's en de vakinstellingen. Het betreft standaardisering van het beroepsgerichte deel van het onderwijs, niet van het generieke deel (dat bestaat uit taal, rekenen, burgerschap, enz.). De onderhavige maatregel omvat de volgende drie aspecten:

- Het gebruik van zogenoemde *sectorale examenprofielen (SEP)*: dit omvat per sector afspraken over te hanteren examens, de dekking ten opzichte van de dossiers, de kwaliteitsborging en de betrokkenheid van het bedrijfsleven bij de examens. Er zijn nu 24 profielen opgesteld die voor een grote groep van 240 kwalificatiedossiers gelden. Zij omvatten deels generieke eisen voor alle dossiers in het profiel en deels dossier-specifieke eisen. Het format is landelijk bepaald, de afspraken moeten regionaal worden uitgewerkt. De SEP's garanderen de inhoud en dekking van het examen en zorgen voor kwaliteitsborging en voor betrokkenheid van het bedrijfsleven.
- Het stimuleren van het *centraal inkopen* van examens in plaats van het zelf ontwikkelen daarvan. De inkoop kan plaatsvinden bij de kenniscentra – die al van oudsher een rol vervullen in de examenontwikkeling – bij uitgeverij of bij particuliere bureaus. De inspectie voert de controle uit op de kwaliteit daarvan. Op dit moment wordt 50% aan centrale examens ingekocht; het is de bedoeling dat dit 60% wordt.
- Het stimuleren van *samenwerking* tussen instellingen, zodat er meer gebruik wordt gemaakt van elders reeds ontwikkelde examens. Dit lijkt vooral voor de kleinere opleidingen van belang.

Naast inkoop bestaat er voor sommige opleidingen een mogelijkheid van landelijke afname van examens, maar dat zijn uitzonderingen. Een voorbeeld zijn opleidingen voor banketbakers, die ovens nodig hebben, en opleidingen voor de scheepvaart, waar met simulaties moet worden gewerkt.

Het betreft geen wettelijke maatregel, maar het gaat om afspraken die binnen de sector zelf worden gemaakt. Het ingezette beleidsinstrument heeft – na een aanvankelijk subsidie – het karakter van stimuleren en overtuigen.

8.3 Doel

Het ultieme doel van de maatregel ‘standaardisering van beroepsgerichte examens in het mbo’ is het borgen van de kwaliteit van het mbo-examen. Er moet meer eenduidigheid komen (standaardisatie) in de beroepsgerichte examinering door het gebruik van examenprofielen. Het bedrijfsleven moet meer betrokken zijn bij de examinering. Daarnaast moet de kwaliteit van de examens dusdanig in orde zijn dat het bedrijfsleven ook meer vertrouwen heeft in de diploma's. Verder beoogt de maatregel meer doelmatigheid

in de examinering te realiseren door centrale inkoop en onderlinge uitwisselingen en samenwerking. Ten slotte wordt ernaar gestreefd de transparantie en vergelijkbaarheid van examens te vergroten. De maatregel is een succes als meer examens van mbo-opleidingen van voldoende kwaliteit zijn.

8.4 Veronderstellingen over gedragsreacties en effecten

De huidige wijze van examinering in het middelbaar beroepsonderwijs heeft niet altijd het gewenste niveau. De algemene veronderstelling van deze maatregel is dat door het hanteren van profielen, het benutten van centraal ontwikkelde examens of het uitruilen en samenwerken op het gebied van examinering in de sector betere examens voor de beroepsgerichte vakken komen. Onderzoek liet zien dat standaardisering vele voordelen kent, en het beleid is daarop gebaseerd. De veronderstelling is dat instellingen/opleidingen met gestandaardiseerde examens beter scoren in het inspectieoordeel, en dat de verstrekte diploma's op meer vertrouwen in het bedrijfsleven kunnen rekenen. Wat betreft het samenwerken lijkt de aanname te zijn dat het samen een examen ontwikkelen een vorm van controle door *peers* is, waardoor een zekere kwaliteitstoets wordt ingebouwd.

Effecten op instellingsniveau

De maatregel zal instellingen meer bewust maken van de kwaliteit van examens. Instellingen zullen, hoewel het geen wettelijk vereiste is, zich toch gaan inzetten voor standaardisering van de examens. Zij kunnen het zich immers niet permitteren om een negatief oordeel van de inspectie te krijgen en zo in de publiciteit te komen. Standaardisering maakt het mogelijk om instellingen te vergelijken, en instellingen willen zo hoog mogelijk in een eventuele *ranking* komen. Een andere reden voor instellingen om zelf (met de regionale arbeidsmarkt) aan de slag te gaan met standaardisering van de examens is eveneens een negatieve: de angst voor centrale examinering als de sector zelf niet in standaardisering slaagt.

Het is nog wel de vraag hoe scherp instellingen – na invoering van standaardisering – zullen zijn op het handhaven van de sectorprofielen.

Effecten op docentniveau

Wellicht heeft het – op termijn – verplichtende karakter van de profielen het meeste effect op een gedragsverandering bij de docenten. Maar het is voorstelbaar dat men nog niet goed in staat is daarnaar te handelen en vooralsnog slechts kleine wijzigingen aanbrengt. Het is de vraag welke prikkel de docenten het meest aanspoort: is dat een negatieve (de angst voor licentie-intrekking bijvoorbeeld) of juist een positieve?

Effecten op studentniveau

De mbo-studenten zullen met gestandaardiseerde examens en de daaraan verbonden instemming van het bedrijfsleven een diploma in handen krijgen dat ook echt iets voorstelt. Wel zal de maatregel in sommige gevallen tot zwaardere examens, en daarmee tot meer selectie leiden. De zwakkere studenten zullen dus studievertraging oplopen dan

wel af- of voortijdig uitstromen. De betere studenten hebben echter baat bij de maatregel, omdat het diploma meer erkende waarde heeft.

Onbedoelde effecten en beleidsafwijkingen

Het is de vraag of alle instellingen/opleidingen zich aan de maatregel willen committeren. Voor recent slecht presterende opleidingen lijkt het meer voor de hand te liggen om de examens te kunnen standaardiseren dan voor de al goed presterende opleidingen/instellingen. In dat geval wordt het doel van onderlinge vergelijkbaarheid niet gehaald: de inspectie zal immers niet de profielen zelf meenemen in het toezichtkader. Een ander gevaar doet zich voor bij de kleinere opleidingen waarvoor gestandaardiseerde examens niet of heel moeilijk te ontwikkelen zijn. Deze zullen toch aangewezen blijven op zelf ontwikkeld materiaal.

8.5 Evaluatie

Bij de evaluatie staat de vraag centraal of het beleid van standaardisering leidt tot een betere kwaliteit van de beroepsgerichte examens. De rol van de overheid is gelegen in het stimuleren van samenwerking tussen instellingen, het bevorderen van inkoop van examens en het introduceren van sectorale examenprofielen. De verwachting is dat hiermee de kwaliteit van de beroepsgerichte examens zal verbeteren.

De centrale variabele in de effectevaluatie is de kwaliteit van de beroepsgerichte examens. Bron daarvoor is het oordeel van de onderwijsinspectie.

Effectevaluatie

De centrale maatstaf voor de effectevaluatie is de kwaliteit van de examens: het beleid is effectief als het percentage beroepsgerichte examens dat voldoet aan de kwaliteitseisen toeneemt. Het bevorderen van samenwerking tussen instellingen en van inkoop van examens en het werken met examenprofielen zijn middelen om dit doel te bereiken. Er is weliswaar een streefcijfer voor het percentage ingekochte examens, maar het halen van dat streefcijfer is geen indicatie dat het doel van het beleid wordt gerealiseerd.

Het beleid om de kwaliteit van beroepsgerichte examens te bevorderen is integraal beleid. De effectevaluatie kan dus niet uitgaan van de vergelijking tussen een experimentele groep die het beleid ondergaat en een – al dan niet equivalente – controlegroep die er niet aan onderworpen is. Wel zijn er mogelijk verschillen binnen het mbo tussen opleidingen of sectoren waar de kwaliteit van de examens al vergaand op orde is en andere opleidingen of sectoren waar dat nog veel minder het geval is. Wellicht zijn informatiebronnen over de examenkwaliteit te vinden bij de Inspectie van het Onderwijs, de Kenniscentra Beroepsonderwijs Bedrijfsleven en de Centra voor Innovatief Vakmanschap. De effectevaluatie zou kunnen aansluiten bij dergelijke verschillen: nagegaan kan worden of het verschil in examenkwaliteit tussen zwakke en sterke opleidingen of sectoren de komende jaren vermindert. De sterke opleidingen of sectoren fungeren hierbij als controlegroep waarvoor het beleid geen grote impact heeft, de zwakkere vormen de experimentele groep. Als uit herhaalde metingen blijkt dat de verschillen in

examenkwaliteit tussen beide groepen afnemen, is dat een indicatie voor de impact van het beleid. In een verkennend vooronderzoek zou moeten worden nagegaan of er in het mbo verschillen in examenkwaliteit zijn, die een aangrijpingspunt kunnen vormen voor deze aanpak.

Deze opzet gaat uit van variaties in beleidsintensiteit. De veronderstelling is dat het beleid in sectoren waar de examens al vaak op orde zijn weinig impact heeft en weinig aanleiding is om in actie te komen. Als zou blijken dat ook in sterke sectoren veel veranderingen rondom de beroepsgerichte examens op gang komen, gaat die veronderstelling niet op. Sterke sectoren kunnen dan niet meer worden beschouwd als de controlegroep, waar weinig wordt gedaan met het beleid. De metingen krijgen dan het karakter van een monitor die laat zien of de doelstelling van het beleid wordt gerealiseerd, maar die geen uitspraken over effecten van het beleid mogelijk maakt.

Een andere optie voor de effectevaluatie is het benutten van mogelijke faseverschillen in de invoering van standaardisering van examens. Opleidingen of sectoren die eerder beginnen met standaardisering fungeren daarbij als experimentele groep, sectoren die later van start gaan zijn de controlegroep. Uit herhaalde metingen kan blijken of de trend in beide groepen verschilt.

In de opleidingen voor de zorgsector is men met het oog op de beroepspraktijk al vrij ver met standaardisering van beroepsgerichte examens. Het lijkt daardoor te laat om de zorgsector nog als experimentele groep te hanteren. Een voormeting van examenkwaliteit is hier niet meer mogelijk. Wel zou kunnen worden nagegaan of er eerdere gegevens beschikbaar zijn over de kwaliteit van examens die als voormeting in deze en andere sectoren kunnen fungeren.

Omdat er nog veel onduidelijkheden bestaan, zou het bijzonder nuttig zijn een vooronderzoek te verrichten naar de mogelijkheden van een quasi-experimenteel design.

Verklarende evaluatie

De verklarende evaluatie is bedoeld om de mechanismen te achterhalen waardoor een maatregel in de praktijk wel of niet (geheel) werkt. Onderzocht zou moeten worden welke rol de ingezette middelen spelen. Is de verbetering van de kwaliteit van de examens toe te schrijven aan vormen van standaardisering (inkoop van examens, samenwerking tussen instellingen, afstemmen van examens op het sectorale examenprofiel)?

Als eerste stap zou de veronderstelling moeten worden getoetst dat gestandaardiseerde examens van betere kwaliteit zijn dan examens die decentraal binnen de opleidingen zijn ontwikkeld. Een kwestie die hierbij speelt, is dat de beleidstheorie niet volledig helder is. Zo wordt verondersteld dat centraal ingekochte examens beter zijn, maar de inspectie heeft in een aantal gevallen geconstateerd dat dit niet het geval was. In deze evaluatie moet die veronderstelling dus eerst uitgebreider worden getoetst. Dat kan door vergelijking van instellingen/opleidingen die wel en die niet gebruikmaken van de centrale inkoop.

In een tweede stap kan vervolgens worden onderzocht of een verbetering van de kwaliteit van examens bij bepaalde opleidingen of in bepaalde sectoren te herleiden is tot een toename van de standaardisering van de examens in die sector.

De verklarende analyse moet ook helder maken wat de stimuli of motiverende factoren zijn voor instellingen om goed inhoud te (blijven) geven aan gestandaardiseerde examens. Is dat de dreiging van de inspectie, angst voor het imago, kosteneffectiviteit of de angst voor centrale examinering? Wat zijn de mechanismen waardoor instellingen alert blijven op de examenkwaliteit en wat zijn factoren waardoor de aandacht verslapt? Wat drijft de docent? Een combinatie van kwantitatief en kwalitatief onderzoek kan helder maken hoe het onderwijsveld reageert op de maatregel, en antwoorden bieden op deze vragen.

Een aanvullend onderzoek, waarbij cases (instellingen/opleidingen) die nog niet met de profielen werken systematisch worden vergeleken met cases waar men is overgegaan op de profielen en daarbij behorende examens, moet de verdere benodigde informatie opleveren. Leiden deze betere examenprofielen ook echt tot betere examens? Hoe kijken gebruikers tegen de nieuwe standaarden aan en wat denken niet-gebruikers te missen – of missen ze niets? Wat is het oordeel van de inspectie, van studentenorganisaties en van stakeholders uit het bedrijfsleven? Wat kunnen we leren van de goede voorbeelden, welke mechanismen zorgen voor succes en zijn deze te vertalen naar andere situaties/cases?

9 Versteving van de kennisbasis in het hbo middels landelijke toetsen

9.1 Achtergrond en aanleiding

De achtergrond van deze maatregel is het gebrek aan kwaliteit van alternatieve afstudeertrajecten zoals dat vorig jaar (2011) aan het licht kwam bij verschillende opleidingen van InHolland en later ook bij andere instellingen in het hoger onderwijs. Naar aanleiding van deze kwestie wordt in het beleid sterker ingezet op externe validering van opleidingen en examens en op versterking van de kennisbasis bij studenten in het hoger beroepsonderwijs.

Aanvankelijk was het beleidsstreven om in het hoger beroepsonderwijs centraal te examineren. Dit is echter gezien de grote diversiteit aan opleidingen niet goed mogelijk. Hooguit is het mogelijk om op enkele kernvakken centraal te toetsen. Bij universitaire studies als medicijnen is men daar bijvoorbeeld al vrij ver mee en worden studenten middels een landelijke set van diagnostische toetsen een paar maal per jaar getest op hun kennisbasis. Dergelijke toetsen maken het mogelijk om de kennistoename bij studenten te meten, en zeggen ook iets over de kwaliteit van het onderwijs dat ze krijgen.

In deze aanzet staat *versterking van de kennisbasis in het hoger beroepsonderwijs via gezamenlijke toetsontwikkeling en landelijke toetsing op kernvakken* centraal. De maatregel wordt niet geïsoleerd toegepast, maar maakt deel uit van een set van beleidsmaatregelen die de kwaliteit van het hoger onderwijs moeten garanderen. Er zijn plannen om het accreditatiekader aan te scherpen. De Inspectie van het Onderwijs krijgt een grotere rol en in het hoger onderwijs wordt risicogericht toezicht ingevoerd. Verder zijn maatregelen aangekondigd om de onafhankelijkheid van de examencommissie te vergroten en wordt een aanwijzingsbevoegdheid voor de minister gecreëerd. Naast aanpassingen die een wetswijziging vergen, zijn door de NVAO al verschillende veranderingen doorgevoerd die de accreditatiepraktijk versterken. In de Strategische Agenda, het Hoofdlijnenakkoord en prestatieafspraken werd al duidelijk dat er meer wordt ingezet op de kwaliteit van docenten (meer masteropgeleide docenten), voldoende contacturen en meer profilering in het onderwijsaanbod (OCW 2012). De hoofdlijnenakkoorden vormen de basis voor de prestatieafspraken. De voorgestelde maatregel past kortom binnen een beleid dat gericht is op een structurele verbetering van de kwaliteit van het hoger onderwijs.

9.2 Inhoud

Deze maatregel verkeert in de beginfase van het beleid. Feitelijk betreft het hier niet een eenduidige maatregel die voor alle opleidingen in het hoger onderwijs zonder meer is toe te passen. Centrale toetsontwikkeling is een complex en langdurig proces. Kern van 'de maatregel' is dan ook het stimuleren van meer samenwerking in het hoger beroepsonderwijs op het gebied van toetsing, zodat in de toekomst per sector meer gezamenlijke toetsing op kernvakken zal plaatsvinden. Dat kan overigens niet voor alle

opleidingen en bij alle vakken (denk aan dansopleidingen). Maar bij verschillende technische opleidingen zal het bijvoorbeeld wel mogelijk zijn om landelijke wiskundetoetsen te ontwikkelen en af te nemen. Het is voorstelbaar dat alle studenten voldoende kennis van wiskunde moeten hebben om deze opleidingen succesvol te kunnen volgen en later te kunnen functioneren op de arbeidsmarkt. Ook in de lerarenopleidingen is op dit vlak ervaring opgedaan: hier is een toets per type lerarenopleiding ontwikkeld. Gezien de complexiteit en de duur van het ontwikkeltraject van centrale toetsen zal zoveel mogelijk worden aangesloten bij bestaande initiatieven, zoals die van SURF¹, de instellingen en de HBO-raad. Een van die initiatieven betreft bijvoorbeeld een toets-itembank voor statistiekexamens, maar er zijn ook toetsen voor verpleegkunde, wiskunde, en economie.

Het beleid verkeert nog in een beginfase. Momenteel is een commissie bezig met het opstellen van een advies over de kenniscomponenten die deel zouden moeten uitmaken van landelijke toetsing. Het betreft kenniscomponenten van kernvakken in zeven verschillende sectoren van het hoger beroepsonderwijs. Het is nog onduidelijk op welke termijn de eerste centrale toetsen in enkele of alle zeven sectoren van het hoger beroepsonderwijs daadwerkelijk kunnen worden afgenomen. Voor dit beleid is 8 miljoen euro uitgetrokken.

9.3 Doel

Het doel van de maatregel *versterking van de kennisbasis in het hoger beroepsonderwijs via gezamenlijke toetsontwikkeling en landelijke toetsing op kernvakken* is de introductie van een kwaliteitsminimum in het hoger beroepsonderwijs, waardoor diploma's ook echt wat gaan voorstellen. De maatregel wil een zekere kennisbasis per opleiding garanderen, maar impliciet ook een bewustwordingsproces starten bij instellingen over het niveau van examinering. Instellingen moeten zich er meer van bewust worden dat zij verantwoordelijk zijn voor de waarde van het diploma van hun studenten. Doelen van de maatregel zijn dus: een verhoging van het kennisniveau van afgestudeerden in het hoger onderwijs op kernvakken, het verbeteren van de kwaliteit van examens en het borgen van de waarde van de diploma's. De maatregel is een succes als er geen onregelmatigheden meer voorkomen op het vlak van diplomaverstrekking en er sprake is van handhaving van het niveau.

9.4 Veronderstellingen over gedragsreacties en effecten

De huidige wijze van examinering in het hoger beroepsonderwijs heeft niet altijd het gewenste niveau. De algemene veronderstelling is dat door samenwerking tussen instellingen in de sector betere toetsen (in kernvakken) van voldoende niveau voorhanden komen en men daarvan ook zal gebruikmaken. De aanname lijkt daarbij te zijn dat samenwerking in het ontwikkelen hiervan een vorm van controle door peers inhoudt en men de lat dus niet te laag zal leggen.

Effecten op instellingsniveau

Instellingen zullen zich door deze beleidsmaatregel, maar ook door aanpalend beleid, meer bewust worden van de kwaliteit van de door hen aangeboden opleidingen en het bijbehorend eindniveau. Voor instellingen die dit signaal goed oppakken, zal het een reden zijn om de interne toezichtstructuur aan te pakken, zodat deze goed functioneert en blijft functioneren. De verwachting is ook dat examencommissies een sterkere rol gaan vervullen dan tegenwoordig het geval is.

Hoewel het hier in eerste instantie examinering in kernvakken en vooraansnog in zeven sectoren in het hoger beroepsonderwijs betreft, is het denkbaar dat er in het gehele hoger beroepsonderwijs een discussie ontstaat over wat een hbo-waardig diploma is. In die zin zou er een uitstralingseffect van de maatregel uitgaan. De vraag is hoe ver dat reikt en hoe lang het zal aanhouden.

Effecten op docentniveau

Landelijke examinering stelt eisen aan de docenten. De verwachting is dat deze hun lesaanbod zullen aanpassen om hun studenten beter voor te bereiden op de landelijke toetsen.

Effecten op studentniveau

Zwak presterende studenten zullen meer worden uitgedaagd zich in te spannen voor hun studie en zich de vereiste basiskennis eigen te maken. Dat zal tot selectie leiden, maar geeft de studenten die het niveau wel aankunnen de garantie van een waardevast diploma, op niveau en maatschappelijk onomstreden.

Onbedoelde effecten en beleidsafwijkingen

Een beleidsbeperking is dat de maatregel niet alle vakken en ook niet alle opleidingen (tegelijk) kan beslaan. Daarmee blijft de invloed op de waarde van diploma's in het hoger beroepsonderwijs mogelijk beperkt. Landelijke toetsen op kernvakken zeggen wel iets, maar niet alles over de kwaliteit van een opleiding, zeker als die goeddeels uit andere, meer specifieke vakken bestaat.

Het is denkbaar dat instellingen/docenten de zwaardere landelijke toetsen op kernvakken gaan compenseren op onderdelen waarvoor (nog) geen landelijke toetsen beschikbaar zijn. Aan die keuze kunnen verschillende drijfveren ten grondslag liggen (het niveau is al hoog genoeg, op die manier slaagt er niemand meer, die landelijke toetsen zeggen ook niet alles, wat was er mis met onze oude toetsen, we worden niet rijk van een groot aantal gezakte studenten en het doet onze naam geen goed).

Aangezien de maatregel sterk afhangt van enkele initiatieven en door het veld zelf (mede-)ontwikkeld worden, is het denkbaar dat het beleid niet of niet goed van de grond komt.

Een andere kwestie is dat het hier om een langdurend proces gaat, waardoor de kans dat in de sector op een te laag niveau wordt geëxamineerd nog geruime tijd kan blijven bestaan.

Weer een ander probleem doet zich voor als het draagvlak binnen instellingen te beperkt is. Dan kiest men wellicht voor een ‘cosmetische aanpassing’ van afstudeertrajecten, maar zal men niet toestaan dat er minder studenten slagen voor een opleiding. Om de maatregel te laten slagen, zal eigenaarschap moeten optreden onder alle actoren in het hoger beroepsonderwijs: van docenten tot (financieel) management en examencommissies. Mogelijk ontstaat een strijd tussen de docenten die deze kernvakken geven en de overige aan een opleiding verbonden docenten. Het is de vraag op welk niveau de zwaktes zitten: bij een vakdocent die een student niet wil frustreren in diens studie of bijvoorbeeld bij een financieel management dat is gebaat bij uiteindelijk een hoog percentage geslaagden. Mogelijk lukt het de doorgaans zeer grote instellingen niet om op alle niveaus draagvlak voor dit doel te creëren.

Het lijkt onvermijdelijk dat een strengere toetsing negatief uitpakt voor sommige zwak presterende studenten en studenten die boven hun niveau hebben gereikt. Deze studenten zullen studievertraging oplopen, uitvallen of afstromen naar een lager niveau. Betreft het echt grote aantallen, dan wordt het voor instellingen moeilijker om een breed en gevarieerd aanbod overeind te houden.

9.5 Evaluatie

Effectevaluatie

Dit beleid verkeert in een heel pril stadium. Op dit moment is nog onduidelijk welke kenniscomponenten van welke studies in aanmerking komen voor landelijke toetsen. Ook is onduidelijk welke toetsen nu reeds landelijk beschikbaar zijn en welke nog ontwikkeld zullen moeten worden.

Verder hangt veel af van de vraag of aan het beschikbaar komen van landelijke toetsen ook een verplichting verbonden wordt om deze te gebruiken. Als dat zo is, dan is de maatregel voor alle instellingen/opleidingen integraal geldig en kunnen geen experimentele groep en controlegroep worden afgebakend. Van een robuuste effectmeting kan in dat geval dus geen sprake zijn. Is het echter zo dat instellingen/opleidingen facultatief kunnen gebruikmaken van landelijk ontwikkelde toetsen, dan zijn er wel mogelijkheden. In dat geval zouden instellingen kunnen worden gevraagd mee te doen aan een experiment waarbij studenten volledig random worden opgesplitst in een groep die deelneemt aan de nieuwe landelijke toetsen, waarop het onderwijsprogramma is afgestemd, en een groep die deelneemt aan de toetsen volgens het oude regiem. Als beide groepen volledig vergelijkbaar zijn, kan een gevonden verschil in prestaties worden toegeschreven aan een verschil in toetsing – een eerste signaal dat de pogingen om de kennisbasis te versterken het gewenste effect hebben gehad. De toetsen moeten inhoudelijk wel op elkaar worden afgestemd in een externe validering.

Deelname aan het experiment zou als subsidievoorwaarde kunnen worden opgenomen. Een mogelijke uitkomstmaat is de toekomstige loonontwikkeling van beide groepen studenten.

Het ligt voor de hand ontwikkelingen in het hoger beroepsonderwijs goed te monitoren. Daarbij zou moeten worden gelet op ontwikkelingen in het kennisniveau op de kernvakken en op rendementgegevens: gekwalificeerde uitstroom, aantal gezakten, afstromers, studiewisselingen en de arbeidsmarktpositie van afgestudeerden. Belangrijk is ook om te onderzoeken hoe de verhouding tussen prestaties op landelijke toetsen en de op andere wijze getoetste prestaties in andere vakken/onderdelen zich per student EN per opleiding ontwikkelen. Is het nu zo dat er stelselmatig veel lager gescoord wordt op de landelijke toetsen of liggen de scores hier in lijn met de overige prestaties van studenten?

Daarnaast zijn tevredenheidsgegevens van docenten, van studenten tijdens en na hun studie en van werkgevers relevant. Wel is het zo dat de resultaten op grond van vergelijkingen in de tijd niet mogen worden toegeschreven aan de maatregel. Vele andere factoren waarvoor niet gecontroleerd kan worden, zijn namelijk eveneens van invloed op de uitkomsten.

Verklarende evaluatie

Dit type onderzoek is bedoeld om te achterhalen door welke mechanismen een maatregel wel of niet goed uitpakt. Het ligt voor de hand om per instelling allereerst te kijken hoe men reageert op de maatregel. 'Men' is hier breed en omvat zoals eerder gezegd vele actoren: van bestuur en intern toezicht tot examencommissieleden en docenten.

Het is belangrijk om meer in het algemeen de kenmerken en *triggers* te achterhalen waardoor sommige instellingen zich meer bekommeren om het handhaven van een basaal kwaliteitsminimum dan andere. Wat zijn de kenmerken van instellingen/opleidingen en docenten die veel waarde hechten aan niveauhandhaving en wat zijn de kenmerken van instellingen/opleidingen en docenten die kiezen voor een 'flexibele of pragmatische' aanpak?

Dat zijn echter meer algemene onderzoeksvragen, die weliswaar zeer relevant zijn, maar nog niet direct betrekking hebben op de maatregel van landelijke toetsen. Een belangrijke vraag die zich daarbij opdringt, is wat het effect van landelijke toetsing in de ogen van betrokkenen is. Leidt dit tot niveauhandhaving en hbo-waardige diploma's? En wat is het effect van landelijke toetsing afgezet tegen meer imagogerichte factoren als negatieve publiciteit, *naming and shaming* in de media?

Daarnaast is onderzoek relevant naar het *draagvlak* in de instellingen en bij verschillende actoren (ook studenten zelf) voor meer landelijke toetsing in het hoger onderwijs teneinde een zekere kennisbasis te garanderen. Het is belangrijk een analyse te verrichten die de verschillende actorperspectieven laat zien.

Naast vragen naar het draagvlak zijn vragen aan de betrokkenen over de verwachte en de waargenomen effecten van de maatregel relevant. Deze vragen zouden ex ante en ex post kunnen worden voorgelegd.

Voor het laten welslagen van de maatregel zou in een verklarend onderzoek tevens ruime aandacht moeten uitgaan naar het inventariseren van knelpunten bij de implementatie. Zoals we hierboven hebben geschetst, kunnen die zich op verschillende vlakken en bij verschillende actoren voordoen. Door een dergelijk onderzoek snel na invulling van de

maatregel te starten en er daarmee een ex durante beleidsevaluatie van te maken, kunnen opgedane bevindingen nog leiden tot aanpassingen van de maatregel.

Noot

- 1 SURF werd in de jaren tachtig van de vorige eeuw opgericht door de gezamenlijke Nederlandse universiteiten, de HBO-raad en de onderzoeksinstellingen. Het doel was om samen te werken aan ICT-innovaties op universiteiten, hogescholen en onderzoeksinstellingen. Inmiddels heeft SURF zijn werkveld verbreed tot onder andere landelijke samenwerking en toetsontwikkeling.

10 Aanscherpen toezicht door de inspectie in het hoger onderwijs

10.1 Achtergrond en aanleiding

In het voorjaar van 2011 kwam aan het licht dat hogeschool InHolland bij vier verschillende opleidingen onrechtmatig diploma's had verstrekt aan studenten via zogenoemde alternatieve afstudeertrajecten. Onderzoek van de inspectie wees uit dat bij andere hogescholen eveneens verschillende opleidingen 'zeer zwak' of 'voor verbetering vatbaar' waren. Ook daarna kwamen er verschillende voorbeelden in de publiciteit van onterecht verstrekte diploma's. De problemen lijken zich vooral voor te doen in het hbo en veel minder in het wetenschappelijk onderwijs. Mogelijke verklaringen daarvoor zijn dat de kwaliteit van het management in het hbo vaker te wensen overlaat, dat docenten lager opgeleid zijn en dat de controle door peers geringer is dan in het wetenschappelijk onderwijs. Recent waren er echter ook signalen van enkele onrechtmatigheden op universiteiten. Zo berichtten de media dat het tentamenniveau regelmatig naar beneden wordt bijgesteld, dat scripties van een te gering niveau toch met voldoende worden gewaardeerd en dat er te gemakkelijk 'genadezesjes' worden uitgedeeld.

De staatssecretaris heeft vorig jaar in reactie op de negatieve berichten in een brief aan de Tweede Kamer verschillende maatregelen aangekondigd om de kwaliteit van het hoger onderwijs en de te behalen diploma's / het eindniveau te garanderen. Naast specifiek op de frauderende instellingen en instellingen met 'zorgelijke opleidingen' gerichte maatregelen zijn dat ook maatregelen op stelselniveau. Een daarvan staat hier centraal, namelijk een meer systematische controle op kwaliteit van de Inspectie van het Onderwijs van opleidingen in het hoger onderwijs. Die rol is nieuw. Tot nu toe hield de inspectie zich wel bezig met de kwaliteitsborging, maar niet met de geleverde kwaliteit in het hoger onderwijs. Dat is immers de taak van de Nederlands-Vlaamse Accreditatieorganisatie (NVAO). Deze bezoekt de opleidingen echter maar één keer in de zes jaar. Dat wordt nu te weinig gevonden: men wil 'niet alleen de foto, maar ook de film'. De inspectie is daarom in overleg met het veld bezig met het ontwikkelen van een toetsingskader voor het hoger onderwijs. De inspectiecontroles zullen plaatsvinden tussen de accreditatierondes van de NVAO in. Daarmee verandert de verhouding tussen de inspectie en de NVAO.

10.2 Inhoud

De Inspectie van het Onderwijs gaat anders dan in het verleden toezicht houden op de kwaliteit van het hoger onderwijs, met name op het eindniveau. De visitaties van opleidingen vinden landelijk per cluster plaats, zodat opleidingen met elkaar kunnen worden vergeleken. De visitatiepanels worden voortaan benoemd door de NVAO. Het toezicht van de inspectie is risicogericht en proportioneel, dat wil zeggen: op basis van signalen en andere risico-elementen zal de inspectie optreden. Indicatief zijn rendementsgegevens – waarbij zowel hoge als zeer lage rendementen de aandacht vragen – dalende

instroomcijfers of grote schommelingen in studentenaantallen, de tevredenheid van studenten en medewerkers, de omvang en spreiding van opleidingen over vestigingen of bijvoorbeeld de 'leeftijd' van een opleiding (ocw, beleidsreactie 2011). De inspectie zal handelen op basis van de beschikbare gegevens en deze niet zelf verzamelen. De rapporten van de inspectie worden openbaar. In het uiterste geval kan de inspectie de accreditatie overdoen, en eventueel kan zij deze intrekken. In het lichtste geval wordt een verbetertraject ingezet en valt de opleiding onder de eerstvolgende accreditatie. De nieuwe rol van de inspectie wordt geregeld in de Wet op het onderwijstoezicht (WOT). Het wetsvoorstel komt nog deze zomer (2012) in de Tweede Kamer. De maatregel zou dit najaar (2012) al van kracht kunnen worden.

10.3 Doel

Het bezoek van accreditatiecommissies eens in de zes jaar heeft bij het management van wo- en hbo-instellingen niet geleid tot een scherp bewustzijn over de kwaliteit van opleidingen en met name het behaalde eindniveau. De verwachting is dat de inspectie dit bewustzijn door toezicht tussen de accreditatiebezoeken in zal versterken. Het is de bedoeling dat de inspectie aanvullend op het werk van de NVAO problemen in het hoger onderwijs tijdig boven tafel krijgt, zodat deze kunnen worden aangepakt. De effecten van het toezicht van de inspectie op het hoger onderwijs moeten zichtbaar worden in de kwaliteit van de onderwijsprocessen, met name in de wijze van examinering en in de onderwijsprestaties.

De doelen van de maatregel zijn het verbeteren en het waarborgen van de kwaliteit van het hoger onderwijs, de wijze van examinering en het gerealiseerde eindniveau. Kernpunt van de borging van de onderwijskwaliteit is gelegen in de externe validering van de toetsing en de examinering. De maatregel kan een succes worden genoemd als schandalen als dat rond InHolland voortaan uitblijven, instellingen hun wijze van examinering op orde hebben en in de maatschappij niet getwijfeld wordt/hoef te worden aan het niveau van een wo-/hbo-opleiding of de waarde van een wo-/hbo-diploma. De ultieme uitkomstmaat betreft deze externe legitimatie van het diploma.

10.4 Veronderstellingen over gedragsreacties en effecten

Incidenten in het hbo en het wo die duidelijk maken dat de kwaliteit van diploma's of tentamens niet (altijd) gegarandeerd is, tasten het maatschappelijk vertrouwen aan in de waarde van de diploma's. Dit risico is groter naarmate het aantal excessen toeneemt. Daardoor kan zelfs de indruk ontstaan dat het niet om incidenten gaat, maar dat er structureel iets mis is met de kwaliteit van het hoger onderwijs.

Het feit dat de inspectie in andere onderwijssectoren toezicht uitoefent, heeft niet kunnen voorkomen dat er (zeer) zwakke scholen of risicoscholen ontstaan. Wel lijkt onderzoek naar het effect van inspectiebezoeken in het primair onderwijs erop te wijzen dat de prestaties in het onderwijs daardoor omhooggaan. Dit effect zou zich sterker voordoen naarmate de bezoeken intensiever zijn (Luginbuhl et al. 2007; Janssens en De

Wolf 2007). Onomstotelijk bewijs van een causaal effect van het toezicht op de kwaliteit van het onderwijs ontbreekt echter.

De algemene beleidstheorie gaat ervan uit dat het effect van risicogericht toezicht kwaliteitsverbetering is. Door vaker te controleren en (eerder) te reageren op signalen kan voorkomen worden dat onderwijsinstellingen slechte opleidingen of een ondermaatse examinering handhaven. Naast een controlerende en signalerende functie gaat er naar verwachting ook een preventieve werking van extra (naast accreditatie) toezicht uit. De algemene veronderstelling van de maatregel is dat voldoende en systematische controle op de kwaliteit van het hoger onderwijs en de verstrekte diploma's schandalen als InHolland zal voorkomen, en maatschappelijke twijfel over de waarde van afgegeven diploma's zal wegnemen. Impliciet wordt daarmee ook gesteld dat het huidige systeem van externe borgingsmechanismen en interne kwaliteitszorg door instellingen zelf faalt.

Een van de problemen is de financieringswijze in het hoger onderwijs. Deze leidt tot perverse prikkels, al is dit effect onlangs sterk gereduceerd. Om al te gretige diplomaverstrekking te voorkomen let de inspectie al langer op de samenstelling van examencommissies: daarvan mogen financiële medewerkers geen deel uitmaken.

Effecten op instellingsniveau

Het effect van risicogericht en proportioneel toezicht van de inspectie is dat het management van instellingen in het hoger onderwijs zich meer bewust wordt van kwaliteitsborging en op dit vlak ook maatregelen zal nemen. Het betekent dat op de verschillende lagen in de organisatie – bestuurders, interne toezichthouders en ook docenten – een bewustwordingsproces op gang moet komen om de interne kwaliteitscontrole serieus aan te pakken. Feedback van de inspectie moet de instellingen met problemen daarbij helpen, zodat zij gericht de noodzakelijke acties zullen gaan ondernemen.

Effecten op studentniveau

Studenten zijn het meest gebaat bij een diploma dat iets voorstelt. Uiteraard zijn zwakke studenten wel de dupe van deze maatregel, als alle docenten ertoe overgaan om minder makkelijk genadezesjes voor tentamens en scripties te geven en instellingen de examinering serieuzer nemen en minder makkelijk diploma's verstrekken.

Onbedoelde effecten en beleidsafwijkingen

Voor de inspectie is deze rol in het hoger onderwijs nieuw. Mogelijk dat de inspectie onvoldoende is toegerust om deze goed te kunnen vervullen. In dat geval zullen effecten uitblijven of pas na langere tijd optreden.

Ook zouden zich problemen kunnen voordoen bij de samenwerking tussen de NVAO en de inspectie. Een mogelijk probleem is rolvervaging tussen beide partijen.

Instellingen zullen als gevolg van het verscherpte toezicht mogelijk minder geneigd zijn te experimenteren met nieuwe opleidingen. Juist nieuwe opleidingen lopen meer risico

door startproblemen. Risicomijdende gedrag van instellingen kan ook noodzakelijke vernieuwing in de weg staan.

10.5 Evaluatie

Effectevaluatie

In de 'Leidraad effectmeting bij inspecties' wordt onder een effect verstaan: 'een waarneembare verandering in hetgeen met de toezicht- en handhavingsactiviteiten van de inspectie wordt beoogd, als gevolg van de activiteiten van de inspectie' (Vossen 2005). Onderzoek op dit vlak is bedoeld om aan te tonen dat de activiteiten die de inspectie onderneemt inderdaad leiden tot een waarneembare verandering in het gedrag van instellingen rond de kwaliteitsborging. Onderzoek naar effecten en neveneffecten van toezicht is echter complex en dit onderzoeksgebied is nog tamelijk onontgonnen (Janssens en De Wolf 2007). Om een causaal effect te kunnen aantonen is een experimentele opzet of een quasi-experimenteel design noodzakelijk. Het al eerder genoemde onderzoek van Luginbuhl et al. (2007) heeft via een natuurlijk experiment wel (beperkt) kunnen vaststellen dat door (frequent) inspectiebezoek de prestaties van leerlingen op de Cito-toets verbeterden.

Aangezien de nieuwe taak voor de inspectie gevolgen heeft voor het hele hoger onderwijs op het moment dat de wet van kracht wordt, kan in het onderhavige geval niet aan de voorwaarde van een experimentele groep en controlegroep worden voldaan. Het toezicht is ook niet random: de inspectie bezoekt opleidingen/instellingen alleen als er een reden is om dat te doen.

Indien er 'wachlijsten' zouden ontstaan van opleidingen die toezicht behoeven, is de vorming van een experimentele groep en een controlegroep wellicht mogelijk. Als de inspectie random (bv. naar regio) opleidingen bezoekt uit die grotere te onderzoeken groep, dan vormen de bezochte opleidingen de experimentele groep en de opleidingen op de wachtlijst de controlegroep. Met een difference-in-differencebenadering kan worden onderzocht of in de experimentele groep de examens verbeteren door het bezoek en de feedback van de inspectie. Dit is echter een hypothetische optie, aangezien de kans dat zich zulke wachlijsten zullen voordoen niet erg waarschijnlijk is.

In dat geval laat de maatregel zich waarschijnlijk het best evalueren door ontwikkelingen in het hoger onderwijs goed te monitoren. Ook dan is het de vraag of instellingen/opleidingen die zijn bezocht door de inspectie zich ook verbeteren. Het beantwoorden van die vraag op basis van monitoring levert belangrijke informatie op, maar een causaal verband met het inspectiebezoek is niet te leggen aangezien een controlegroep ontbreekt. Het is belangrijk voor ogen te houden dat de bij monitoring waargenomen veranderingen niet het resultaat van het beleid hoeven te zijn. Daarvoor spelen te veel andere factoren eveneens een rol.

Verklarende evaluatie

Dit type evaluatie is bedoeld om de mechanismen te achterhalen waardoor risicogericht toezicht van de inspectie wel, niet of niet onder alle omstandigheden werkt. Het design van de verklarende evaluatie hangt (mede) af van de meest relevante beleidsvragen.

De algemene onderzoeksvragen die met deze verklarende evaluatie beantwoord zouden moeten worden, zijn: Ontstaat binnen de hoger onderwijs meer kwaliteitsbewustzijn door het toezicht van de inspectie, en waarin uit dat verhoogde kwaliteitsbewustzijn zich (bij welke instellingen, opleidingen of actoren en op welk vlak)? Wordt het horizontale toezicht gestimuleerd door meer verticale verantwoording en leidt het tot meer transparantie? Zijn de instrumenten die de inspectie inzet (alle even) geschikt?

Daarnaast zijn er meer specifieke vragen te formuleren naar de aard van het toezicht en naar de (re-)acties van instellingen – zowel de instellingen waarvan een of meer opleidingen onder toezicht staan als de overige instellingen/opleidingen voor hoger onderwijs. Vragen over *het toezicht* zijn bijvoorbeeld: Hoeveel instellingen en opleidingen bezoekt de inspectie jaarlijks en op welke gronden gebeurt dat? Is het bezoek dekkend voor de problematiek? Ofwel: kunnen alle opleidingen waarmee iets aan de hand lijkt te zijn ook door de inspectie worden bezocht of zullen er opleidingen op een wachtlijst komen? Welke signalen en andere risico-elementen zijn het meest bepalend voor daadwerkelijk gesignaleerde tekortkomingen of misstanden? Hoe verlopen de samenwerking en de communicatie tussen de NVAO en de inspectie? Hoe beoordelen beide partijen elkaars interventies? Welke *checks and balances* zijn voor een instelling noodzakelijk om de kwaliteit van opleidingen en het diploma te kunnen borgen?

Vragen over de *instellingen/opleidingen* waarmee iets aan de hand is, zijn: Hoe verlopen de samenwerking en de communicatie tussen inspectie en instelling? Kunnen de instellingen zich vinden in het oordeel van de inspectie? Zijn de instellingen in staat om de feedback van de inspectie te vertalen in een handelingsplan? Hoe traag of hoe snel zijn de instellingen in staat om veranderingen aan te brengen bij opleidingen die tekortschieten? Beslaat de aanpassing/hervorming niet te veel tijd in de ogen van betrokkenen (onder wie studenten)? Welke omstandigheden beïnvloeden de kwaliteitsborging in positieve en negatieve zin?

Er zijn verschillende mogelijkheden voor onderzoek op basis van deze vragen. Een meerjarige opzet van zowel kwantitatief als kwalitatief onderzoek ligt voor de hand. De wet moet echter nog worden aangenomen en het beleid moet nog worden uitgevoerd. Effecten worden mogelijk pas na enkele jaren zichtbaar. Het evaluatieonderzoek heeft het karakter van een *ex durante* evaluatie. Ervaringen die worden opgedaan tijdens het onderzoek kunnen leiden tot aanscherping van de maatregel of verbetering van het beleidsproces.

11 Kwaliteit/prestatieafspraken en profiel in de bekostiging

11.1 Achtergrond en aanleiding

In de strategische agenda voor het hoger onderwijs 'Kwaliteit in verscheidenheid' worden diverse maatregelen aangekondigd om de kwaliteit van het hoger onderwijs te verhogen en de verscheidenheid te bevorderen. Het kabinet bouwt hiermee voort op de aanbevelingen uit het rapport van de commissie-Veerman 'Differentiëren in drievoud'. Een van de maatregelen die worden aangekondigd, is een aanpassing van de bekostigingssystematiek om keuzes in het profiel van instellingen te faciliteren en instellingen te belonen die een hoge kwaliteit leveren. Dit wordt aangeduid als 'kwaliteit en profiel in de bekostiging'. De maatregelen die onder deze noemer worden getroffen, komen in de plaats van eerdere collectieve meerjarenafspraken op brancheniveau met de sectoren hbo en wetenschappelijk onderwijs. Op basis van deze afspraken is tot op heden onvoldoende voortgang geboekt, reden om nu over te gaan op afspraken per afzonderlijke instelling.

11.2 Inhoud

De overheid maakt met individuele instellingen in het hbo en het wetenschappelijk onderwijs afspraken over te leveren kwaliteit en over profilering van de instelling. Daarvoor komt in het bekostigingsmodel van het hoger onderwijs een onderdeel 'kwaliteit en profiel', met een omvang van 7% van de onderwijsbekostiging. De middelen die eerder beschikbaar waren in het kader van de afspraken op brancheniveau waren een aanvulling op de reguliere bekostiging. Deze middelen worden nu ook ingezet voor het deel 'kwaliteit en profiel' in de bekostiging. De middelen zullen onderdeel zijn van de reguliere bekostiging. Het bedrag dat is gereserveerd voor 'kwaliteit en profiel' loopt op tot 310 miljoen euro in 2015. Het grootste deel van het bedrag (5% van de 7%) is bestemd voor prestatieafspraken (voorwaardelijke financiering). De prestatieafpraak betreft de ambities van een instelling op het gebied kwaliteit, profilering en valorisatie. Het tekenen van een prestatieafpraak is voorwaardelijk voor het ontvangen van de 5% van de bekostiging vanaf 2013. Een kleiner deel van het geld (2% van 7%) wordt selectief toegekend voor profilering.

De *prestatieafspraken* hebben betrekking op studiesucces en onderwijskwaliteit. Bij studiesucces betreft het voorkomen van uitval in het eerste jaar en het rendement van de bachelorfase. Onderwijskwaliteit heeft met name betrekking op meer excellentie binnen de opleidingen, maar kan ook bredere kwaliteitsindicatoren betreffen (studentenoordelen over kwaliteit). Daarnaast worden onder de noemer maatregelen afspraken gemaakt over de docentkwaliteit, de onderwijsintensiteit en de hoeveelheid indirecte kosten (overhead) binnen de instelling. Voor elk van deze aspecten zijn

standaardindicatoren geformuleerd, maar instellingen die dat willen, kunnen ook alternatieve indicatoren aandragen waarvoor ze doelstellingen formuleren.

De middelen voor de prestatieafspraken worden toegekend voor een periode van vier jaar. Aan het einde van die periode worden de inspanningen en prestaties van de instellingen beoordeeld. Voorwaarde voor opnieuw toekennen van de middelen voor een periode van vier jaar is dat alle prestaties op het onderdeel kwaliteit en studiesucces worden behaald. Als dat niet zo is, kan een instelling worden gekort (er wordt na vier jaar niet afgerekend op de onderdelen profilering en valorisatie, deze onderwerpen maken echter wel onderdeel uit van de prestatieafpraak). De verwachting is dat met alle instellingen prestatieafspraken zullen worden gemaakt.

Het onderdeel *profilering* betreft een aanscherping van het profiel van hogescholen en universiteiten op het gebied onderwijs en onderzoek. Via een selectieve toedeling van middelen worden instellingen aangemoedigd om een eigen profiel te ontwikkelen. In het hoger beroepsonderwijs kan het eigen profiel betrekking hebben op publiek-private samenwerkingsverbanden gericht op topsectoren of maatschappelijke sectoren zoals zorg en onderwijs, op professionele masters en *associate degree*-programma's en op *honours*-trajecten en driejarige programma's voor vwo-ers. In het wetenschappelijk onderwijs gaat het om het bevorderen van zwaartepuntvorming in onderwijs en onderzoek en om verdere differentiatie van het onderwijs.

Een reviewcommissie beoordeelt de profileringsvoorstellen en adviseert over de voorwaardelijke financiering en de toekenning van de selectieve middelen voor profilering. De middelen worden toegekend voor een periode van vier jaar, met halverwege een midterm-review. De verwachting is dat niet alle profileringsvoorstellen zullen worden gehonoreerd. Het betreft een selectieve toedeling waarvan niet alle instellingen zullen profiteren. Na vier jaar volgt een nieuwe ronde met prestatieafspraken en toekenning van selectieve middelen.

Op de langere termijn zal het aandeel in de bekostiging van 'kwaliteit en profiel' – nu 7% – mogelijk verder omhooggaan. Dit is afhankelijk van de ervaringen die de komende jaren met het beleid worden opgedaan en van de beschikbaarheid van (extra) middelen.

11.3 Doel

De doelstellingen van het beleid 'kwaliteit en profiel in de bekostiging' zijn:

- verbetering van het studiesucces in het hbo en wetenschappelijk onderwijs;
- verhoging van de kwaliteit van het onderwijs en meer differentiatie van het onderwijs (naar niveau en naar inhoud);
- versterking van de profilering van de instellingen zodat zwaartepunten en differentiatie in onderwijsprogramma's ontstaan;
- het ontstaan van een sterkere band tussen onderwijs en onderzoek;
- meer aandacht voor kennisvalorisatie.

Intermediaire doelstellingen met het oog op verbetering van het studiesucces en verhoging van de onderwijskwaliteit zijn:

- verhoging van de kwaliteit van docenten;
- vergroting van de onderwijsintensiteit;
- het binnen perken houden/terugdringen van de indirecte kosten.

Door betere docenten, meer contacturen en minder overhead zullen het studiesucces en de onderwijskwaliteit toenemen, zo is de gedachte.

Studiesucces en onderwijskwaliteit zijn op het niveau van de afzonderlijke instellingen af te lezen. Profilering en het ontstaan van zwaartepunten hebben betrekking op de verhouding tussen het aanbod van verschillende instellingen: doen alle instellingen hetzelfde of heeft elke instelling een eigen profiel?

11.4 Veronderstellingen over gedragsreacties en effecten

Het beleid 'kwaliteit en profiel in de bekostiging' gaat uit van de werkzaamheid van financiële prikkels. Voor het onderdeel 'kwaliteit' maakt de overheid afspraken met instellingen over streefcijfers op het gebied van studiesucces en kwaliteit, met daaraan gekoppeld een bonus/malus-regeling. De 5% wordt alleen toegekend als inderdaad een prestatieafpraak wordt afgesloten in 2012. Bij het niet halen van de doelen krijgen instellingen in de volgende periode minder budget. Gekozen is voor financiële consequenties op de iets langere termijn. Bij het onderdeel 'profiel' moeten instellingen met elkaar concurreren om schaarse middelen. Het gaat om een tamelijk gering deel van de bekostiging (2%), waar wellicht een niet heel sterke prikkel van uitgaat. Niettemin signaleert o c w al verschillende initiatieven bij instellingen.

Effecten op instellingsniveau

De gemaakte afspraken en de daaraan gekoppelde financiële prikkel zullen instellingen aanzetten tot acties gericht op verhoging van het studiesucces. Te denken valt aan meer keuzebegeleiding en zorgvuldiger intake van nieuwe studenten, en misschien ook het stellen van strengere eisen aan de geschiktheid en motivatie van aanstaande studenten. Instellingen zullen ook maatregelen moeten nemen om te bevorderen dat studenten vlotter studeren (bewaking van de studievoortgang, studiebegeleiding). Sommige instellingen hanteren sinds kort al veel strengere voortgangseisen dan tot dusverre gebruikelijk was. Een strengere selectie op geschiktheid bij de toegang tot opleidingen zou ertoe kunnen leiden dat opleidingen minder studenten zullen toelaten en dus kleiner zullen worden.

Om te voldoen aan de doelstelling van meer excellentie zullen instellingen het aanbod voor goede studenten moeten uitbreiden (bv. via honoursprogramma's). Er zal een meer divers aanbod aan opleidingen ontstaan.

Personeel aan de universiteiten geeft niet alleen onderwijs, maar doet ook onderzoek. De gewenste intensivering van het onderwijs zou kunnen leiden tot een zekere verschuiving in takenpakket van onderzoek naar onderwijs. De druk om goed te presteren op het

gebied van onderzoek is echter groot, en aanzienlijke verschuivingen zijn waarschijnlijk niet te verwachten.

De profilering en de vorming van zwaartepunten zullen kunnen leiden tot een verandering in het onderwijsaanbod van instellingen. Als instellingen zich toeleggen op bepaalde zwaartepunten, zullen zij andere onderdelen mogelijk afstoten. Het aanbod van een instelling kan daardoor smaller worden. De verwachting is ook dat de keuze voor een bepaald profiel leidt tot meer kwaliteit van het onderwijs dat de instelling aanbiedt.

Effecten op studentniveau

De prestatieafspraken kunnen voor studenten verschillende gevolgen hebben. Mogelijk krijgen ze te maken met een strengere toelating tot opleidingen, waarbij ze meer dan voorheen worden beoordeeld op geschiktheid en motivatie. Na hun toelating zullen studenten waarschijnlijker harder moeten werken om te voldoen aan de strengere voortgangseisen die instellingen mogelijk zullen gaan hanteren. Van de langstudeerdersmaatregel gaat overigens een vergelijkbare invloed uit. Studenten die in aanmerking willen komen voor excellentietrajecten zullen meer inspanningen moeten leveren. In een stelsel met meer differentiatie tussen instellingen zal de bewust kiezende student wellicht vaker uitwijken naar een verder weg gelegen instelling om daar de gewenste kwaliteit of aanpak te vinden.

Effecten op stelselniveau.

Als de prestatieafspraken worden gehaald nemen de kwaliteit van het onderwijs en de efficiëntie van het stelsel toe (verhouding ingezette middelen en behaalde diploma's). Doordat instellingen een eigen profiel kiezen waarmee ze zich onderscheiden van andere instellingen wordt de differentiatie op stelselniveau groter. Door specialisatie neemt ook de kwaliteit op stelselniveau mogelijk toe.

Onbedoelde effecten en beleidsafwijkingen

Het bevorderen van de toegankelijkheid van opleidingen, het vergroten van het rendement en het verhogen van de kwaliteit van het onderwijs zijn lastig tegelijkertijd te realiseren. Een vergroting van het rendement kan ten koste gaan van de toegankelijkheid, omdat strenger wordt geselecteerd op kans op studiesucces, of van de kwaliteit, omdat daaraan concessies worden gedaan om een hoog rendement te halen. In het huidige beleid lijkt te worden gekozen voor rendement en kwaliteit. Dat kan tot gevolg hebben dat de toegankelijkheid van opleidingen onder druk komt te staan. Om de rendementsdoelstellingen te halen zonder concessies te doen aan de kwaliteit, zullen instellingen studenten mogelijk strenger gaan selecteren op geschiktheid.

Specialisatie en zwaartepuntvorming hebben niet de gewenste uitwerking als studenten hun studiekeuze baseren op overwegingen van bereikbaarheid, bijvoorbeeld onder druk van de versraling van studiefinanciering. De gewenste match tussen instellingen met een bepaald profiel en studenten komt dan niet tot stand.

11.5 Evaluatie

Bij effectevaluatie draait het om de vraag naar de relatie tussen de interventie en de uitkomsten. Nemen het studiesucces en de kwaliteit van het onderwijs toe en is dat te danken aan de prestatieafspraken die met de instellingen worden gemaakt? Neemt de profilering van instellingen toe en is dat het gevolg van de selectieve toedeling van middelen voor dat doel?

De verklarende evaluatie betreft de vraag naar de mechanismen die verantwoordelijk zijn voor het optreden (of achterwege blijven) van effecten. Hoe worden de maatregelen uitgevoerd en kunnen de gevonden effecten daaruit worden verklaard?

Effectevaluatie

Of de doelen van het beleid worden gehaald, kan voor het onderdeel *onderwijskwaliteit en studiesucces* worden afgelezen aan de indicatoren waarover afspraken met instellingen zijn gemaakt. Uit het halen van de doelen kan echter niet worden geconcludeerd dat de prestatieafspraken daadwerkelijk effect hebben gehad. Om dat te kunnen vaststellen moet een vergelijking worden gemaakt tussen een groep die aan de maatregel is blootgesteld en een vergelijkingsgroep waarop het beleid niet van toepassing is. Omdat het de bedoeling is met alle instellingen prestatieafspraken over het studiesucces te maken, is dat echter niet mogelijk.

Het maken van een vergelijking tussen opeenvolgende instroomcohorten – ingestroomd respectievelijk voor en na de afspraken met de instellingen – biedt geen soelaas. Het studiesucces wordt namelijk ook beïnvloed door andere veranderingen in het hoger onderwijs, zoals de introductie van de langstudeerdersmaatregel. Volstaan zal moeten worden met het monitoren van doelstellingen voor studiesucces en onderwijskwaliteit. De effectmaten voor het monitoren van het onderdeel ‘studiesucces en kwaliteit’ liggen vast in de afspraken met de instellingen.

Bij het onderdeel *profilering* kan waarschijnlijk wel onderscheid worden gemaakt tussen instellingen die middelen krijgen om zich te profileren en instellingen die hiervoor geen middelen krijgen. Alle instellingen kunnen een aanvraag indienen voor extra middelen, maar naar verwachting zullen niet alle voorstellen worden gehonoreerd. Er kan dus een vergelijking worden gemaakt tussen instellingen waarvan de aanvraag is gehonoreerd en instellingen waarvan de aanvraag is afgewezen. Robuuste uitspraken over effecten van de toekenning van middelen voor profilering kunnen op basis van zo'n vergelijking echter niet worden gedaan. Omdat de beste voorstellen worden gehonoreerd en de slechtste worden afgewezen, zal de controlegroep niet goed vergelijkbaar zijn met de experimentele groep. Als er een rangordening van instellingen is te maken op basis van de kwaliteit van de ingediende voorstellen kan worden nagegaan of het vergelijken van instellingen wier voorstel net wel en instellingen wier voorstel net niet is gehonoreerd, is een werkbare optie is. Een complicatie is dat het aantal afgewezen voorstellen mogelijk niet zo groot zal zijn, waardoor de controlegroep van instellingen met afgewezen voorstellen een te geringe omvang heeft.

Als een effectevaluatie niet mogelijk is, zal ook voor het beleidsonderdeel ‘profilering’ moeten worden volstaan met monitoren. Hoewel dan geen harde uitspraken over effecten kunnen worden gedaan, zou daarbij wel aandacht moeten worden besteed aan verschillen tussen instellingen met profileringsbudget en instellingen zonder dat budget.

Waarom kan de mate profilering van een instelling worden afgemeten? Een mogelijke maatstaf voor de mate van profilering van een instelling is het aantal studenten dat kiest voor die instelling vanwege het specifieke profiel. Dat kan worden gemeten via een peiling bij studenten. Een andere mogelijkheid is dat bij relevante partijen (bv. docenten in het hoger onderwijs, schooldecanen voortgezet onderwijs, andere betrokkenen) gepeild wordt hoe zij aankijken tegen het profiel van verschillende instellingen voor hoger onderwijs. Ook zou kunnen worden overwogen om de mate van profilering van instellingen kwalitatief te laten beoordelen door deskundigen. Daarnaast kan worden onderzocht of de indicatoren die zijn ontwikkeld voor de ranking van instellingen aanknopingspunten bieden voor het meten van de mate van profilering.

Verklarende evaluatie

Voor de prestatieafspraken over studiesucces en onderwijskwaliteit is een effectevaluatie in strikte zin niet haalbaar. Niettemin is het zinvol om te onderzoeken wat instellingen doen om de gestelde doelen te halen. Via onderzoek bij hogescholen en universiteiten zou kunnen worden nagegaan welke maatregelen instellingen treffen om het studierendement te verbeteren en wat de ervaringen met de gekozen aanpak zijn. Eerder is gesignaleerd dat instellingen verschillende accenten kunnen leggen. Ze kunnen bijvoorbeeld strenger gaan selecteren ‘aan de poort’, de studiekeuzebegeleiding en intake intensiveren, de loopbaanbegeleiding tijdens de studie intensiveren en het onderwijs zelf intensiveren.

De maatregelen die instellingen nemen, zullen deels al naar voren komen uit de prestatieafspraken die instellingen maken en uit de verantwoording die zij jaarlijks afleggen over deze afspraken. Op basis hiervan – zo nodig aangevuld met een beknopte peiling onder instellingen voor hoger onderwijs – kan worden nagegaan welke mix van maatregelen de instellingen treffen. Door de instellingen op grond van hun keuzes in te delen in verschillende typen, kan een verantwoorde selectie van cases worden gemaakt. In een vergelijkend casestudyonderzoek kan vervolgens worden bezien met welke mix van maatregelen – en in welke context – de beste resultaten worden gehaald. Door per instelling systematisch meerdere relevante actoren over bovengenoemde vragen te raadplegen, wordt het inzicht in werkzame principes vergroot – met dien verstande dat geen harde uitspraken over effectiviteit van verschillende strategieën mogelijk zijn.

Een relevante vraag voor het verklarend onderzoek is ook welke rol de financiële prikkel speelt bij de prestatieafspraken. Stimuleert de prikkel de instellingen tot actie en passen ze hun beleid aan of worden de middelen ingezet voor bestaand beleid rondom studiesucces en onderwijskwaliteit?

De centrale vraag wat betreft de profilering van instellingen is hoe deze de praktijk haar beslag krijgt en welke effecten zij heeft voor de positie van de instelling. Een kwalitatief

onderzoek onder instellingen kan daar licht op werpen. Nagegaan zou kunnen worden op welke wijze de profilering tot stand is gekomen: met welke andere instellingen en partijen wordt samengewerkt? Wat houdt het profiel in? Gaat de profilering gepaard met het afstoten van onderdelen van het opleidingsaanbod van een instelling naar andere instellingen? Ervaart men de financiële prikkel als een stimulans of zou de profilering naar het oordeel van de betrokkenen ook zonder prikkel tot stand zijn gekomen? Welke rol speelt excellentie in de profilering van instellingen? Heeft men de ervaring dat het profiel bekend is en studenten aanspreekt, en heeft het profiel wervingskracht? In een kwalitatief onderzoek kunnen zowel instellingen worden betrokken die budget hebben gehad voor profilering als instellingen aan wie hiervoor geen budget is toegelikt. Vergelijking tussen beide typen instellingen kan duidelijk maken welke rol de extra middelen spelen in het instellingsbeleid op het terrein van profilering.

12 Collegegelddifferentiatie

12.1 Achtergrond en aanleiding

Het hoger onderwijs kent tot op heden een brede toegankelijkheid en nauwelijks enige vormen van selectie. Selectie – anders dan door loting – gebeurde alleen bij de kunst-academies, de dansacademies, de academies voor lichamelijke opvoeding en in het kader van numerus-fixusopleidingen.

Het kabinet wil excellentie bevorderen in alle sectoren van het onderwijs en dus ook in het hoger onderwijs. Hiertoe zou er ruimte moeten komen voor het selecteren van studenten voor opleidingen. Zo'n selectie van studenten krijgt dan betere onderwijs-programma's aangeboden en, omdat men aanneemt dat betere opleidingen ook duurder zijn, mag daarvoor een hoger collegegeld worden gevraagd.

Met selectie van studenten en collegegelddifferentiatie wordt al geëxperimenteerd. In het kader van Ruim Baan voor Talent (RBVT) vonden verschillende experimenten plaats die excellentie beoogden te bevorderen. De aanbeveling destijds van de Commissie ruim baan voor talent was dat om een cultuur van excellentie en diversiteit te creëren binnen een instelling sprake moet zijn van enige massa in deelname, zodat het excellentiebeleid zichtbaar is en impact heeft op de gehele instelling. Bekend van honoursprogramma's is dat zich hiervoor doorgaans maar weinig deelnemers aanmelden. Men is er onbekend mee, terwijl de programma's toch al enige tijd bestaan. In het voortgezet onderwijs worden leerlingen er op geattendeerd en de arbeidsmarkt vraagt er (nog) niet om. Wel krijgt ook het voortgezet onderwijs nu extra geld om excellentie te stimuleren en de verwachting is dat er op termijn een keten van excellent onderwijs ontstaat. Het doel van het al lopende honoursprogramma Sirius¹ is te achterhalen op welke manier excellentie het beste kan worden bereikt en hoe dit het beste is aan te sturen.

12.2 Inhoud

De strategische agenda voor het hoger onderwijs bevat verschillende maatregelen die gericht zijn op het stimuleren van excellentie. De onderhavige maatregel staat er als volgt beschreven: 'Instellingen mogen voor duurdere, excellente, trajecten meer collegegeld vragen. Zij hebben dan meer mogelijkheden om zich op excellentie te profileren. Via collegegelddifferentiatie kunnen extra middelen aangeboord worden, in aanvulling op de rijksbijdrage. De meerinkomsten kunnen worden benut voor extra voorzieningen om een hoger niveau te bereiken en vast te houden, zoals extra docenten, gastdocenten of extra studiebelasting. De kosten voor deze voorzieningen hoeven dan niet ten laste te komen van het budget voor de reguliere opleidingen. Tegen deze achtergrond wordt collegegelddifferentiatie via het wetsvoorstel Ruim Baan voor Talent mogelijk gemaakt bij opleidingen met *residentieel, kleinschalig en intensief onderwijs* (o.a. de university colleges).¹ Aanvullend op deze optie bestaan de volgende mogelijkheden:

- *programma's gericht op een hoger eindniveau*: het gaat hier om *een track* binnen de opleiding, de rest van de opleiding blijft toegankelijk tegen het wettelijk collegegeld;

- *opleidingen met het oordeel excellent* van de Nederlands-Vlaamse Accreditatieorganisatie (NVAO).

‘Hierbij gaat het per definitie om opleidingen met een grote toegevoegde waarde. De koppeling met het oordeel van de NVAO maakt transparant dat de student extra kwaliteit krijgt voor het hogere collegegeld’ (Kwaliteit in verscheidenheid. Strategische Agenda Hoger Onderwijs, Onderzoek en Wetenschap 2011: 20-21).

Hoewel er op dit moment ook al mogelijkheden zijn om te differentiëren met collegegeld, betreft het hier een nieuwe beleidsmaatregel. Het beleid sluit aan bij het eerdere beleid uit Ruim Baan voor Talent, maar is dus wel veranderd. Zo is het campusmodel niet langer noodzakelijk om een meer exclusieve vorm van onderwijs te kunnen aanbieden en hoger collegegeld te vragen.

Het accreditatiestelsel op grond waarvan het predicaat excellent kan worden verleend, is eveneens vernieuwd. In de oude structuur ging het om zeven verschillende facetten en was het mogelijk om op een of meer van deze facetten het predicaat excellent te krijgen. Nu is sprake van een integrale beoordeling. Er zijn drie standaarden waarop een opleiding excellent kan zijn, maar wat telt, is de overallbeoordeling van de NVAO. Een opleiding kan dus niet excellent zijn als de kwaliteit van het onderwijs niet excellent is.

12.3 Doel

Het doel van de maatregel is om het aanbod van excellente opleidingen en onderwijs te verruimen. De verwachting is dat excellente opleidingen en tracks binnen een opleiding duurder zullen zijn. Het verhoogde collegegeld is niet bedoeld als selectiemiddel – selectie moet op andere gronden plaatsvinden. Door de optie van meer collegegeld aan studenten vragen voor meer uitdagende opleidingen zal dat extra aanbod niet ten koste van het reguliere onderwijs hoeven gaan.

Nog te veel talent aan de top blijft onderbenut. Het uiteindelijke doel is een verhoging van het aantal studenten die (kunnen) excelleren in het hoger onderwijs.

12.4 Veronderstellingen over gedragsreacties en effecten

De centrale veronderstelling bij deze maatregel lijkt te zijn dat excellentie onder studenten momenteel nog te weinig wordt bevorderd en dat er dus prikkels (beleidsmaatregelen) nodig zijn om instellingen te stimuleren meer/beter aanbod te creëren voor veelbelovende studenten.

In de strategische agenda is de veronderstelling te lezen dat selectie de motivatie en inzet van de student zal bevorderen, waardoor prestaties omhooggaan en er in het algemeen een uitdagender studiecultuur ontstaat. Het bekende campusmodel van enkele colleges voldoet wel aan de criteria van selectie en excellent aanbod, maar bereikt naar verhouding weinig studenten. Door de mogelijkheid van collegegelddifferentiatie verwacht men dat dit aanbod toeneemt, ook omdat het campusmodel daarvoor niet langer een vereiste is.

Effecten op instellingsniveau

De verwachting is dat instellingen eerder tot het creëren van excellent onderwijs zullen overgaan nu ze ook meer collegegeld van studenten mogen vragen. Op die manier hoeven ze geen middelen te onttrekken aan het reguliere onderwijs. Naar verwachting zullen instellingen zichzelf 'op de kaart willen zetten' met een of meer exclusieve opleidingen of tracks. Als instellingen erin slagen om veel studenten te interesseren voor deze opleidingen, wordt de exclusieve opleiding zichtbaarder en heeft deze een positief uitstralings-effect op de gehele instelling.

Effecten op studentniveau

Veel studenten voelen zich weinig uitgedaagd in het hoger onderwijs, en missen diepgang. De veronderstelling is dat als er een uitdagender aanbod wordt gecreëerd, veelbelovende studenten hiervan ook gebruik zullen maken, waardoor ze beter/excellent gaan presteren. Het hogere collegegeld kan worden bijgeleend, waardoor de extra financiële lasten voor studenten geen (al te grote) hindernis voor deelname hoeft te zijn.

Onbedoelde effecten en beleidsafwijkingen

De belangrijkste ongewenste effecten zijn in het kort: het in gevaar komen van de toegankelijkheid van het onderwijs en een zekere mate van devaluatie van de reguliere opleiding. Er komt meer (zichtbare) differentiatie tussen studenten: studenten die (deels) een excellente opleiding volgen en studenten die regulier hoger onderwijs volgen. In negatieve zin kan men spreken van een tweedeling. Het is voorstelbaar dat niet alle instellingen voorstander zijn van meer selectie en exclusiviteit in het hoger onderwijs. De maatregel kan ook op verzet van studenten stuiten. Instellingen zullen studenten duidelijk moeten kunnen maken welke 'waar' zij in een excellente opleiding of track voor hun hogere collegegeld krijgen. Het is de vraag of instellingen en de beoordeling door de NVAO de studenten op dit punt kunnen overtuigen. Een ongewenst effect zou kunnen zijn dat het verhoogde collegegeld zelf als selectiemechanisme gaat werken. Wellicht zijn bepaalde talentvolle studenten niet bereid om meer te lenen/betalen voor de exclusieve tracks en opleidingen. Het gevaar van zelfselectie bij studenten uit lagere sociaaleconomische milieus is bijvoorbeeld niet ondenkbaar. In zekere zin zou dan een klassenverschil ontstaan.

Voor sommige buitenlandse studenten zal een hoger collegegeld mogelijk nog steeds lager zijn dan het collegegeld in het land van herkomst. Als de tracks en opleidingen hoofdzakelijk door buitenlandse studenten worden gekozen, zal de Nederlandse kennis-economie niet optimaal kunnen profiteren van de maatregel.

Een ander onbedoeld effect zou kunnen zijn dat de meer exclusieve opleidingen en tracks de aantrekkelijkheid van reguliere opleidingen in het hoger onderwijs doen afnemen. De kans daarop lijkt overigens niet heel groot. Wel zou het zo kunnen zijn dat instellingen zich meer willen profileren met hun exclusieve opleidingen dan met de reguliere opleidingen, en daarvoor toch ook extra middelen onttrekken aan het reguliere onderwijs.

Weer een ander ongewenst effect zou kunnen zijn dat instellingen om deze opleidingen te kunnen ‘vullen’ met studenten de normen voor toelating verlagen. De veronderstelde positieve effecten van meer selectie voor excellentie blijven dan uit.

Een van de problemen met de honoursprogramma's is dat deze nog weinig zichtbaar en herkenbaar zijn. Het is de vraag of deze maatregel zo anders is dat dit nu wel zal lukken.

12.5 Evaluatie

De centrale evaluatievraag is tot welke veranderingen de beleidsmaatregel van collegegelddifferentiatie leidt en of deze veranderingen het causale gevolg zijn van de maatregel. Effectief is de maatregel als deze inderdaad excellentie stimuleert, oftewel als het aanbod van opleidingen en het gebruik hiervan door een relatief homogene groep veelbelovende studenten toenemen.

Eerst moet worden bepaald wat de toegevoegde waarde van deze meer exclusieve vorm van onderwijs is. Dat zal echter bijzonder lastig zijn, aangezien dat ook nu al geldt voor het Sirius Programma.

‘De maatregel’ is ook niet meer dan een middel: meer collegegeld voor meer exclusiviteit.

Effectevaluatie

De mogelijkheden voor een echte effectmeting zijn heel beperkt. Wat is de uitkomstmaat? Het is uiteraard eenvoudig om te kijken of er meer aanbod ontstaat en hoeveel studenten daar gebruik van maken. Het uiteindelijke doel is echter om meer uit excellente studenten te halen, en juist dat ‘meer’ is lastig te bepalen.

Het zou eenvoudiger zijn voor de evaluatie als studenten een vaardigheidstoets aflegden die een uitkomstmaat levert, zoals de Amerikaanse GRE², een instroomtoets die moet worden afgelegd omdat de Verenigde Staten geen centraal examen voortgezet onderwijs kennen. Bij de OECD wordt gewerkt aan een pilotonderzoek om te kijken of het vaardigheidsniveau verbetert of niet via het zogenoemde AHELO-project.³ In dat geval zouden centrale eindexamengegevens als ‘voormeting’ kunnen fungeren en de AHELO als nameting, maar zover zijn we nog niet. Studenten die een excellente opleiding hebben gevolgd, zouden dan kunnen worden vergeleken in hun (studie-)loopbaan met studenten die niet aan een prestigieuze opleiding hebben deelgenomen. Dit lost echter niet het probleem van selectiviteit op. De studenten zijn immers niet random over de reguliere en excellente opleidingen verdeeld, maar zijn geselecteerd op basis van prestatieniveau en waarschijnlijk ook motivatie. Daarmee is er geen sprake van een gelijkwaardige controlegroep. Desalniettemin is dit een zinvolle exercitie.

Ook studenten die wel toelating vroegen, maar zijn afgewezen, vormen geen betrouwbare controlegroep. Door de afwijzing zijn ze niet volledig vergelijkbaar met de toegelaten studenten, behalve misschien als er heel veel belangstelling is voor een exclusieve opleiding, maar slechts een beperkt aantal plaatsen beschikbaar is en de toelating op basis van volgorde van aanmelding plaatsvindt. Maar ook dan is een verschil in motivatie niet uit te sluiten. Hooguit zou een ranking van studenten die net wel en net niet

zijn toegelaten, kunnen worden benut, maar ook alleen als er veel belangstelling is en er weinig beschikbare plaatsen zijn.

Het lijkt bovendien niet erg aannemelijk dat het exclusieve aanbod gretige aftrek zal vinden. De honoursprogramma's die er nu zijn, trekken nog geen drommen studenten. Een mogelijke indicator van succes is de mate van toelating tot prestigieuze masters. Krijgen studenten die een als excellent gekwalificeerde track of bacheloropleiding hebben gevolgd eerder toegang tot zulke prestigieuze opleidingen?

Een belangrijke vraag is natuurlijk of het met deze maatregel lukt om daadwerkelijk meer uit studenten te halen door ze meer te bieden, hen meer uit te dagen en ze op een zo hoog mogelijk niveau te brengen. Je zou de kwaliteit van scripties van studenten uit een reguliere en uit een excellente opleiding met elkaar kunnen vergelijken om iets te zeggen over de kwaliteit van het extra aanbod. De studenten zouden ook dan uiteraard wel vergelijkbaar moeten zijn.

Verklarende evaluatie

Het verklarend onderzoek biedt meer mogelijkheden. Het verklarend onderzoek is bedoeld om de mechanismen te achterhalen waardoor de maatregel wel of niet het gewenste effect heeft. Een eerste relevante stap is het monitoren van ontwikkelingen in deelname van studenten aan reguliere versus meer excellente opleidingen, zowel de toegelaten als eventueel afgewezen studenten. De analyse van deze monitorgegevens zou kunnen worden uitgebreid met prestatiegegevens van de studenten (cijfers centraal eindexamen, doorstroomgegevens, tentamencijfers).

Komen er meer excellente opleidingen, die ook (voldoende/meer) studenten weten te trekken en in welke vorm is dat: campus, opleiding of track? De analyses zouden zicht moeten bieden op de verhouding getalenteerde studenten die wel en die niet deelnemen aan de speciale opleidingen, en aldus antwoord bieden op de vraag welke mate van selectiviteit optreedt als gevolg van de maatregel. Aan studenten zou gevraagd kunnen worden naar hun motieven om wel of niet te opteren voor de excellente opleiding. Apart zou ook aandacht moeten worden besteed aan de vraag in hoeverre studenten zich laten leiden door het hogere collegegeld.

Op instellings-/opleidingsniveau is de vraag relevant hoe men studenten (eventueel werft en) selecteert, welk collegegeld men vraagt en waaraan het extra geïnde geld wordt besteed. Gaat het geld gericht naar die opleidingen of gaat het naar iets anders? Daarnaast is het zaak zicht te krijgen op hetgeen instellingen aanbieden in de excellente tracks of opleidingen: een rijker programma, meer contacturen, betere docenten, homogeneren groepen, kleinere groepen, community-vorming of wellicht iets anders. Een vergelijkende casestudy zou meer inzicht kunnen bieden in het succes (of uitblijven daarvan) van een bepaalde aanpak of inrichting van de opleiding of track.

Noten

- 1 Met het Sirius Programma beoogt het ministerie van Onderwijs, Cultuur en Wetenschap kennis en inzicht te verkrijgen in de wijze waarop in het hoger onderwijs de beste studenten tot een zo hoog mogelijk niveau kunnen worden gebracht. Tevens wil het ministerie de belemmeringen identificeren die daarvoor in de huidige situatie bestaan. Met deze regeling wordt aan universiteiten en hogescholen die zich inspinnen om meer excellentie te bereiken in het hoger onderwijs een tegemoetkoming verstrekt.
Het Sirius Programma richt zich op de volledige breedte van het hoger onderwijs. In het kader van dit programma kunnen bekostigde instellingen voor hoger onderwijs alleen of gezamenlijk een aanvraag indienen om meer te halen uit hun beste studenten. Deze aanvragen zijn instellingsbreed en kunnen bestaan uit een palet van maatregelen die de instelling(en) wil(len) inzetten om dit doel te bereiken. Aanvragers geven daarbij zelf aan wat zij onder excellente studenten verstaan en op welke aspecten zij hen willen laten excelleren.
- 2 The Graduate Record Examinations (GRE) is a standardized test that is an admissions requirement for many graduate schools in the United States, in other English-speaking countries and for English-taught graduate and business programs world-wide. Created and administered by Educational Testing Service (ETS) in 1949, the exam aims to measure verbal reasoning, quantitative reasoning, analytical writing and critical thinking skills that have been acquired over a long period of time and that are not related to any specific field of study.
- 3 The Assessment of Higher Education Learning Outcomes (AHELO) will test what students in higher education know and can do upon graduation. More than a ranking, AHELO is a direct evaluation of student performance. It will provide data on the relevance and quality of teaching and learning in higher education. The test aims to be global and valid across diverse cultures, languages and different types of institutions. AHELO will be a tool for:
 - Universities: to assess and improve their teaching.
 - Students: to make better choices in selecting institutions.
 - Policy-makers: to make sure that the considerable amounts spent on higher education are spent well.
 - Employers: to know if the skills of the graduates entering the job market match their needs.

13 School aan zet

13.1 Achtergrond en aanleiding

In de actieplannen ‘Basis voor presteren’ (basisonderwijs) en ‘Beter presteren: opbrengstgericht en ambitieus’ (voortgezet onderwijs) zijn maatregelen aangekondigd om het prestatieniveau in de kernvakken te verhogen en excellentie te bevorderen. Geconstateerd wordt dat het Nederlandse onderwijs internationaal gezien vrij goed presteert, maar dat er indicaties zijn dat het beter kan. De zwakke leerling in het Nederlandse onderwijs doet het weliswaar relatief goed, maar de prestaties van de betere leerlingen blijven achter bij die van de betere leerlingen in andere landen. Daarnaast laat internationaal onderzoek zien dat de prestaties van leerlingen in het Nederlandse basis- en voortgezet onderwijs stagneren of zelfs enigszins achteruitgaan, terwijl die van leerlingen in een aantal andere landen zijn verbeterd. Scholen moeten daarom meer opbrengstgericht gaan werken en meer aandacht gaan besteden aan de beste leerlingen, zodat prestaties over de hele linie verbeteren en er meer uit de beste leerlingen wordt gehaald.

Bestuursakkoorden

In de bestuursakkoorden tussen OCW, de PO-Raad en de VO-raad zijn hierover afspraken gemaakt en maatregelen overeengekomen. In het primair onderwijs gaat het om het versterken van de onderwijskwaliteit in de hele sector (onder meer vermindering van het aantal zeer zwakke scholen en taal- en rekenzwakke scholen), meer opbrengstgericht werken, een beter aanbod voor begaafde leerlingen en professionalisering van leraren en schoolleiders (onder meer door het instellen van een professioneel statuut en een lerarenregister, door het bevorderen van expertise voor opbrengstgericht werken en voor ‘passend onderwijs’ en door een beter HRM-beleid).

De voorgaande thema’s keren in grote lijnen terug in de afspraken met het voortgezet onderwijs. Deze hebben betrekking op het bevorderen van goede prestaties op de kernvakken én brede vorming, meer opbrengstgericht werken, professionalisering van leraren met het oog op het kunnen omgaan met verschillen, beter HRM-beleid en het bevorderen van een ambitieuze leercultuur met aandacht voor excellentie en hoogbegaafdheid.

Toekenning van middelen via prestatiebox

Via de zogenoemde prestatiebox krijgen schoolbesturen in het primair en voortgezet onderwijs extra middelen – in de vorm van een bedrag per leerling – om de doelstellingen uit de bestuursakkoorden te realiseren. Er is binnen de prestatieboxmiddelen geen oormerking van afzonderlijke doelen. Scholen kunnen afhankelijk van hun uitgangssituatie zelf kiezen voor welke doelen uit het akkoord ze de middelen inzetten. In jaarverslagen en via ‘vensters voor verantwoording’ (voortgezet onderwijs) leggen schoolbesturen verantwoording af over hun uitgangssituatie, de gestelde doelen, de

inzet van middelen en de behaalde resultaten. Er is voorzien in het monitoren van de doelstellingen op het niveau van de hele sector (respectievelijk primair en voortgezet onderwijs) met een midterm-review halverwege de looptijd van de akkoorden.

‘School aan zet’

Om scholen en schoolbesturen inhoudelijke steun te geven bij het realiseren van de doelstellingen van de prestatiebox is het programma ‘School aan zet’ in het leven geroepen. ‘School aan zet’ helpt scholen en schoolbesturen desgewenst bij het analyseren van hun uitgangssituatie, het formuleren van hun ambities en het kiezen van een aanpak, en wordt uitgevoerd door het Platform Bèta Techniek.

13.2 Inhoud

Alle scholen (voortgezet onderwijs) en schoolbesturen (primair onderwijs) worden benaderd met het ondersteuningaanbod van ‘School aan zet’. Het is niet verplicht om van dat aanbod gebruik te maken, en het gebruik is geen voorwaarde voor het ontvangen van de prestatieboxmiddelen.

Voorheen was er al een vorm van ondersteuning in het primair en voortgezet onderwijs, maar die functioneerde op basis van subsidieregelingen per afzonderlijk thema. Omdat de inhoudelijke thema’s bij verschillende ondersteunende instanties waren belegd, leidde dit tot versnippering. Met ‘School aan zet’ komt er één loket en krijgen scholen/schoolbesturen ruimte om zelf prioriteiten te stellen binnen de doelstellingen van de prestatiebox. Het ‘School aan zet’-aanbod vervangt de bestaande ondersteuningsstructuur niet, maar heeft een ‘makel- en schakel’-functie tussen scholen/besturen en de bestaande ondersteuningsstructuur.

‘School aan zet’ biedt vijf verschillende expertiseprogramma’s die overeenkomen met de doelen uit de bestuursakkoorden. Het betreft:

- excellentie en hoogbegaafdheid;
- bèta, wetenschap en techniek;
- opbrengstgericht werken en basisvaardigheden taal/lezen en rekenen;
- lerende organisatie en HRM-beleid;
- omgaan met verschillen.

Binnen elk van deze expertiseprogramma worden verschillende activiteiten aangeboden, namelijk:

- ambitiegesprekken met scholen om hun uitgangssituatie en ambities voor de vijf thema’s vast te stellen;
- ‘makel en schakel’-activiteiten (scholen en het aanbod van diensten bij elkaar brengen, transparant maken van het aanbod van diensten);
- twee verdiepingsgesprekken; het eerste is bedoeld om het thema dat de school heeft gekozen te verrijken en om de gekozen doelen nader te specificeren, het tweede is

tevens gericht op het delen van kennis en ervaringen met andere scholen om van elkaar te leren;

- het stimuleren dat – indien nodig – nieuw aanbod tot stand komt in de educatieve markt. Zo nodig kan kortlopend praktijkgericht onderzoek worden uitgezet om effectieve aanpakken op te sporen en vervolgens te verspreiden;
- twee gesprekken om de ‘vinger aan de pols houden’: dat wil zeggen om te reflecteren op het ingezette beleid en om de voortgang op de doelen in kaart te brengen.

Er zal met drie tranches worden gewerkt. De eerste tranche gaat van start in 2012, de tweede in 2013 en de derde in 2014. Omdat het traject voor alle deelnemende scholen afloopt in 2015, heeft het dus een variabele lengte, afhankelijk van de tranche. Bij de derde tranche loopt het programma kort en volgen de gesprekken elkaar snel op.

13.3 Doel

Het ‘School aan zet’-programma is dienstig aan de bestuursakkoorden. De doelstellingen van het programma zijn dus uiteindelijk identiek aan die van de bestuursakkoorden. In het activiteitenplan van ‘School aan zet’ worden deze doelstellingen in grote lijnen overgenomen. Daarnaast zijn doelen geformuleerd voor het bereik van de operatie: het streven is dat 3000 scholen voor primair onderwijs en 450 scholen voor voortgezet onderwijs zich aansluiten bij ‘School aan zet’.

13.4 Veronderstellingen, gedragreacties en effecten

De centrale veronderstelling van ‘School aan zet’ is dat door steun te bieden scholen beter in staat zijn om werk te maken van de doelen uit de bestuursakkoorden, en daardoor uiteindelijk betere resultaten met hun leerlingen zullen halen.

Effecten op instellingsniveau

Scholen en besturen krijgen veel eigen ruimte; deelname aan ‘School aan zet’ is vrijwillig en scholen kunnen zelf kiezen aan welke doelstelling uit de bestuursakkoorden ze willen werken. Een mogelijkheid is dat scholen die het minst ver zijn met beleid voor de doelen uit de bestuursakkoorden het meest geneigd zullen zijn om gebruik te maken van het ondersteuningsaanbod. In de praktijk kan het ook anders werken: misschien melden zich scholen die al het nodige werk maken van onderdelen uit de bestuursakkoorden, en kiezen die scholen voor thema’s waar men al ver mee is. In de praktijk zal moeten blijken hoe scholen gebruikmaken van de keuzevrijheid die ‘School aan zet’ hun biedt.

Door ‘School aan zet’ zullen scholen naar verwachting meer inzicht krijgen in hun uitgangssituatie op de onderdelen waarop de akkoorden zich richten. Daarnaast is de verwachting dat scholen vaker steun zullen zoeken bij deskundigen of bij andere scholen om van elkaars deskundigheid te profiteren. Het streefbeeld van de interventie is een school die opbrengstgericht werkt en die een aanpak heeft voor excellente leerlingen en een doorlopende leerlijn voor wetenschap en techniek. Daarnaast weten de docenten

om te gaan met verschillen tussen leerlingen en heeft de school een lerende organisatie met een goed HRM-beleid.

De bestuursakkoorden lopen tot 2015. Scholen die snel aan de slag gaan, hebben drie jaar de tijd om veranderingen op gang te brengen. ‘School aan zet’ werkt echter met tranches. Voor scholen uit de tweede en met name de derde tranche is de termijn waarop effecten van ‘School aan zet’ zichtbaar moeten worden erg kort. De derde tranche gaat in 2014 van start, een jaar later zouden er al effecten zichtbaar moeten zijn; dat is wel erg snel.

Effecten op leerlingenniveau

De verwachting is dat de veranderingen in aanpak van het onderwijs uiteindelijk leiden tot resultaten bij leerlingen. De akkoorden bevatten een aantal concrete targets op leerlingenniveau, te halen in 2015, met een tussendoel voor 2013. Deze targets hebben betrekking op het prestatieniveau van de 20% best presterende leerlingen, op het eindexamencijfer voor de kernvakken en op het aantal leerlingen dat kiest voor een natuurprofiel of voor exacte vakken in het voortgezet onderwijs.

13.5 Evaluatie

Centrale vragen zijn: draagt ‘School aan zet’ bij aan de realisering van de doelen uit de bestuursakkoorden (effectevaluatie)? En: hoe wordt ‘School aan zet’ ingevoerd en gebruikt door scholen, is de invoering ervan in overeenstemming met de doelen van de operatie en in hoeverre verklaart de wijze van invoering het al dan niet optreden van de effecten van ‘School aan zet’ (verklarende evaluatie)?

Effectevaluatie

Essentieel voor de effectevaluatie is de mogelijkheid een vergelijking te kunnen maken tussen scholen die gebruikmaken van ‘School aan zet’ en scholen die dat niet doen. Omdat naar verwachting niet alle scholen dat zullen doen, is een vergelijking tussen gebruikers en niet-gebruikers mogelijk. Problematisch is echter dat de verdeling over beide groepen scholen niet willekeurig tot stand zal komen. Als gevolg van zelfselectie zullen scholen die gebruikmaken van het aanbod en scholen die dat niet doen mogelijk geen vergelijkbare uitgangssituatie hebben op de doelen uit de bestuursakkoorden. Om zelfselectie uit te schakelen zou een echte experimentele opzet moeten worden gekozen, waarbij een random samengestelde groep scholen en besturen het aanbod van ‘School aan zet’ gebruikt en de rest niet. Dit is echter geen haalbare kaart, omdat alle scholen het aanbod van ‘School aan zet’ krijgen. Er zijn verschillende opties om het probleem van mogelijke vertekeningen door selectiviteit te omzeilen.

Een eerste optie is om te werken met een difference-in-differencebenadering. Bij deze benadering is het geen vereiste dat beide groepen scholen (gebruikers en niet-gebruikers) met elkaar vergelijkbaar zijn, maar de veronderstelling is wel dat de trendmatige ontwikkeling zonder interventie in beide groepen gelijk zal zijn. Onder die

aanname kan een verschil in ontwikkeling tussen de groep die de interventie ('School aan zet') ondergaat en de groep die de interventie niet ondergaat worden toegeschreven aan de interventie.

De aanname van een gelijke trendmatige ontwikkeling is cruciaal, maar kan als zodanig niet worden onderzocht met deze opzet. Door de vergelijking toe te spitsen op scholen die in een aantal relevante opzichten met elkaar vergelijkbaar zijn (bv. naar samenstelling van de leerlingenpopulatie of met eenzelfde kwaliteitsoordeel van de inspectie), kan het risico dat ten onrechte wordt uitgegaan van een gemeenschappelijke trend worden beperkt. Ook wordt in de literatuur aanbevolen om de aanname van een gelijke trend in beide groepen scholen te toetsen met behulp van historische gegevens (HM Treasury 2011: 117). De beschikbaarheid van historische gegevens stelt hieraan echter grenzen. In het voortgezet onderwijs zou de trend in examencijfers voor de kernvakken kunnen worden vergeleken.

Een tweede optie is dat wordt gebruikgemaakt van de gefaseerde uitvoering van 'School aan zet'. 'School aan zet' werkt met drie opeenvolgende tranches. Scholen die in de tweede of de derde tranche instappen, zouden kunnen fungeren als controlegroep voor de scholen die meedoen aan de eerste tranche. Wenselijk is dat de verdeling van scholen over tranches willekeurig tot stand komt, zodat geen selectievertekeningen ontstaan tussen opeenvolgende tranches. Als de scholen met de grootste problemen of de meest gretige scholen in de eerste tranche worden geholpen, is de vergelijkbaarheid tussen tranches niet meer gewaarborgd. Verder is deze aanpak vooral geschikt voor effecten die zich voordoen op niet al te lange termijn, nog voordat de volgende tranche van start gaat. Ten slotte is het van belang dat de aankondiging/het vooruitzicht van de interventie geen gedragsreacties op gang brengt bij scholen in de tweede en derde tranche, waardoor een eventueel effect al wordt gerealiseerd nog voordat de interventie daadwerkelijk van start gaat.

Een derde optie is een variant op de random experimentele benadering. Een zuiver experimentele opzet waarbij een random deel van de scholen de interventie ondergaat en de rest van de scholen niet is – zoals gezegd – niet haalbaar bij 'School aan zet'. Scholen kunnen niet worden gedwongen om gebruik te maken van 'School aan zet' en kunnen er evenmin van worden uitgesloten. Borghans et al. (2009) werken een alternatief uit voor de situatie dat het gebruik van een voorziening voor de hele populatie openstaat, maar waarbij *een verschil in gebruik* van de voorziening in deelpopulaties wordt bewerkstelligd. Hiertoe krijgt een random deel van de populatie het advies gebruik te maken van een voorziening en krijgt de rest van de populatie dat advies niet. In beide random deelpopulaties zijn er gebruikers en niet-gebruikers. In de deelpopulatie die het advies heeft gekregen om gebruik te maken van de voorziening zal het gebruik naar verwachting echter hoger liggen dan in de deelpopulatie die dat advies niet heeft gekregen. Omdat het advies om gebruik te maken van de voorziening at random is verstrekt, zal dit verschil in gebruik in beide deelpopulaties niet het gevolg van zijn van zelfselectie. Vervolgens kan worden nagegaan in welke mate de doelstelling die met de voorziening is beoogd in beide deelpopulaties wordt gerealiseerd. De vergelijking richt zich dus niet op gebruikers

versus niet-gebruikers, maar op de samenhang tussen de mate van gebruik van een voorziening en de mate waarin de doelstelling van de voorziening wordt gerealiseerd.

Voor 'School aan zet' zou dit als volgt kunnen worden uitgewerkt. In aanvang worden alle scholen/schoolbesturen per brief benaderd met het aanbod van 'School aan zet'. In tweede instantie – bijvoorbeeld een maand later – zou een random gekozen deel van de scholen/schoolbesturen nogmaals moeten worden benaderd met een extra wervende uitnodiging om gebruik te maken van 'School aan zet'. Door deze herbenadering zal waarschijnlijk verschil in de mate van gebruik van het aanbod van 'School aan zet' ontstaan tussen de groep die is herbenaderd en de groep die niet is herbenaderd. De effectevaluatie richt zich vervolgens op de mate van doelrealisatie in beide groepen. Voorwaarde is dat de herbenadering daadwerkelijk leidt tot een flink hoger gebruik van 'School aan zet' in de desbetreffende deelpopulatie dan in de deelpopulatie die niet is herbenaderd. In het beleid moet draagvlak zijn voor de keuze om in tweede instantie slechts een deel van de scholen te benaderen.

Elk van deze opties heeft beperkingen. Bij de eerste optie is de veronderstelling van gelijke trend bij de twee onderscheiden groepen van scholen een kritisch punt. Bovendien stellen de beschikbare effectmaten beperkingen bij herhaalde metingen (zie hieronder). Bij de tweede optie is de (aselecte) verdeling van scholen over tranches belangrijk, en daarnaast de termijn waarop de effecten zich voordoen (de aanpak is minder geschikt voor effecten die zich op de wat langere termijn manifesteren). Bij de derde optie is het cruciaal dat het opnieuw benaderen van een deel van de scholen daadwerkelijk leidt tot een substantieel extra gebruik van 'School aan zet'.

Optie drie is aantrekkelijk vanwege de gekozen oplossing voor het probleem van selectiviteit. Een voordeel is daarnaast dat een voor- en een nameting in deze optie in principe niet nodig zijn; één meting volstaat. Te overwegen is om in te zetten op deze optie, maar tegelijkertijd rekening te houden met de andere twee opties. Indien een herbenadering van een random deel van de scholen onvoldoende extra belangstelling voor 'School aan zet' teweegbrengt, kan op de eerste of de tweede optie worden teruggevallen.

In de bestuursakkoorden zijn indicatoren en streefcijfers opgenomen voor de verschillende doelstellingen. Deze doelstellingen en indicatoren zijn in principe de maatstaf waaraan het effect van 'School aan zet' zou moeten worden afgemeten. Voor de effectevaluatie is het een complicatie dat scholen verschillende accenten kunnen leggen binnen de doelstellingen uit de bestuursakkoorden. Eventuele effecten kunnen zich op uiteenlopende terreinen manifesteren en moeten worden gezien in samenhang met de keuze die scholen hebben gemaakt uit het brede palet van doelstellingen.

Verder moet rekening worden gehouden met het gegeven dat veel indicatoren waaraan het realiseren van de doelen in de bestuursakkoorden wordt afgemeten afkomstig zijn uit verschillende onderzoeken onder jaarlijks wisselende steekproeven van scholen (bv. het onderzoek naar kwaliteit van scholen van de onderwijsinspectie, de monitor ouderbetrokkenheid van Ecorys en de Loopbaanmonitor). Trendvergelijking voor die

indicatoren is daardoor alleen mogelijk op groepsniveau (scholen die meedoen aan ‘School aan zet’ versus scholen die dat niet doen), niet op het niveau van individuele scholen. Het aantal scholen in deze steekproeven is niet erg groot, waardoor schattingen weinig precies zullen zijn.

Sommige indicatoren zijn voor een veel groter aantal scholen beschikbaar omdat ze niet uit een beperkte steekproef komen (de Cito-score van de 20% best presterende leerlingen op basisscholen, het eindexamencijfer van de 20% best presterende vwo leerlingen, het eindexamencijfer op de kernvakken in het voortgezet onderwijs, de keuze van exacte profielen). Het is aan te raden om herhaalde metingen te beperken tot deze laatste effectindicatoren.

Het is te overwegen om in het basisonderwijs niet alleen te kijken naar de Cito-score van de 20% best presterende leerlingen, maar ook naar de score van de overige leerlingen. Het bestuursakkoord Primair onderwijs bevat weliswaar geen doelstelling voor de gemiddelde Cito-score, maar het uiteindelijke doel van opbrengstgericht werken is dat de resultaten van alle leerlingen verbeteren. Tegen die achtergrond is het zinvol om ook het gemiddelde prestatieniveau als effectmaat mee te nemen.

Verklarende evaluatie

In de ‘verklarende’ evaluatie staan twee vragen centraal:

- Hoe wordt het programma ingevoerd?
- Kunnen de effecten van het programma – of het niet-optreden van effecten – worden verklaard door de manier waarop het programma is ingevoerd?

Bij de invoering van de ‘School aan zet’ zijn verschillende partijen betrokken: het Platform Bèta Techniek, schoolbesturen (primair onderwijs) en scholen (primair en voortgezet onderwijs). Gegevens verzameld bij de verschillende partijen kunnen een beeld geven van de invoering en verdere uitwerking van het programma.

Een eerste belangrijke vraag is die naar het bereik van ‘School aan zet’: worden de genoemde streefaantallen gehaald (3000 scholen primair onderwijs, 450 scholen voortgezet onderwijs, 150 scholen (voortgezet) speciaal onderwijs)? Het is van belang om inzicht te krijgen in de karakteristieken van de scholen die meedoen aan het programma. Vormen de deelnemende scholen een afspiegeling van het scholenveld of betreft het juist scholen met bepaalde karakteristieken (bv. schoolgrootte of opbrengsten en in het voortgezet onderwijs: vmbo of havo/vwo)? Wat is bij de deelnemende scholen de verdeling over de vijf verschillende expertiseprogramma’s van ‘School aan zet’ en de achterliggende thema’s?

De verklarende evaluatie moet inzicht bieden in de motieven om aan het programma mee te doen en in de motieven voor de keuze voor de verschillende thema’s binnen het programma. Waarom besluiten scholen en schoolbesturen in te gaan op het aanbod van ‘School aan zet’? Wat zijn de overwegingen om voor een bepaald thema te kiezen binnen het programma? Worden deze keuzen vooral ingegeven door de wens om te werken

aan zwakke punten of wil men sterke punten verder uitbouwen? Was er al beleid op de gekozen thema's waarop 'School aan zet' zich richt?

In het primair onderwijs is sprake van een gelaagde benadering: 'School aan zet' benadert in eerste instantie schoolbesturen. Besturen die meerdere scholen onder hun hoede hebben, kunnen zich passief opstellen en het aanbod 'neutraal' doorsluizen naar hun scholen, waarbij het besluit om mee te doen bij de school ligt. Besturen kunnen echter ook een meer actieve opstelling kiezen en op basis van eigen afwegingen bepalen welke van hun scholen ze in contact brengen met 'School aan zet', en voor welk thema. In het primair onderwijs is het van belang dat de evaluatie laat zien hoe deze gelaagde structuur functioneert. Hoe actief en sturend zijn besturen ten opzichte van hun scholen? Zitten besturen en scholen op dezelfde lijn? Een gebrek aan overeenstemming tussen besturen en scholen zal waarschijnlijk afbreuk doen de effectiviteit van de interventie.

Bij de scholen die aan het programma deelnemen, moet de verklarende evaluatie laten zien wat de ervaringen zijn met het programma. Hoe tevreden zijn scholen over de geboden ondersteuning (het ambitie-, het verdiepings- en het reflectiegesprek)? Komen scholen via het programma in aanraking met andere scholen of instanties met relevante deskundigheid? Brengt het programma deelnemende scholen tot een andere aanpak? Hebben scholen het gevoel dat ze dankzij het programma beter zijn toegerust op de gebieden waarop het programma zich richt en/of het thema waarvoor de school heeft gekozen?

Methode van onderzoek

In de verklarende evaluatie moet materiaal worden verzameld bij verschillende partijen. Gegevens over het bereik van het programma zullen beschikbaar zijn bij het Platform Bèta Techniek. In gesprekken met dit platform kunnen bovendien vragen uit de verklarende evaluatie aan de orde komen vanuit het gezichtspunt van de aanbieders van het programma.

In het primair onderwijs moeten gegevens worden verzameld bij (grotere) besturen en scholen. Er kan zowel een kwantitatieve (gestandaardiseerde) als een meer kwalitatieve benadering worden gevolgd. Ook een combinatie van beide benaderingen is nuttig: eerst wordt met een kwalitatieve aanpak verkend hoe de invoering van het programma verloopt en welke factoren een rol spelen en wat de ervaringen zijn met het programma. Daarop aansluitend kan met een kwantitatief onderzoek een representatief beeld worden verkregen.

De verklarende evaluatie moet niet alleen inzicht in de implementatie van het programma, maar ook aanknopingspunten bieden voor de verklaring waarom effecten optreden. Dat gaat het beste als de effectevaluatie en de verklarende evaluatie uitgaan van dezelfde scholen. Als zo'n opzet niet mogelijk is, moet worden uitgegaan van het totaalbeeld dat uit beide evaluaties naar voren komt.

14 Invoering referentieniveaus in het primair, voortgezet en middelbaar beroepsonderwijs

14.1 Aanleiding en achtergrond

Een aantal jaren geleden werd geconstateerd dat de prestaties van leerlingen in het Nederlandse onderwijs onvoldoende op peil zijn. Internationale peilingen wezen op een daling van de prestaties van leerlingen in het basis- en voortgezet onderwijs, de inspectie constateerde dat een deel van de leerlingen met onvoldoende bagage aan het voortgezet onderwijs begint, incidenteel onderzoek in het mbo wees op tekortschietende basisvaardigheden van studenten en ook op de pabo's waren er problemen met het kennisniveau van studenten. Hieruit is de conclusie getrokken dat het niveau van de taal- en rekenvaardigheden van leerlingen en studenten omhoog moet.

De introductie van referentieniveaus voor taal en rekenen is een van de maatregelen die wordt genomen om de leerprestaties te verbeteren. Met deze maatregel volgt het kabinet de aanbevelingen van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008) en van de commissie-Dijsselbloem (TK 2007/2008). De invoering van referentieniveaus hangt samen met en is onderdeel van breder beleid om de taal- en rekenprestaties te verbeteren, zoals de verplichte eindtoets en leerlingvolgsysteem in het primair onderwijs, de invoering van een rekentoets als onderdeel van het centraal examen in het voortgezet onderwijs, de invoering van centrale examens voor taal en rekenen in het mbo en het bevorderen van opbrengstgericht werken.

14.2 Inhoud

Een referentieniveau specificeert een gewenst niveau van vaardigheden op het gebied van taal en rekenen. De verschillende referentieniveaus – vier voor taal en drie voor rekenen – zijn samengebracht in het referentiekader. De referentieniveaus zijn gekoppeld aan schoolsoorten, zodat in de verschillende fasen van de schoolloopbaan en in de verschillende schoolsoorten duidelijk is wat leerlingen en studenten moeten kennen en kunnen op het gebied van taal en rekenen. Op elk referentieniveau is een nader onderscheid tussen een fundamenteel niveau dat door zo veel mogelijk leerlingen moet worden gehaald en een streefniveau dat een uitdagend perspectief biedt voor de betere leerlingen. De bestaande onderwijswetgeving bevatte voor de verschillende sectoren al doelen voor taal en rekenen (kerndoelen basisonderwijs, eindexamenprogramma's voortgezet onderwijs, eindtermen mbo). Door de introductie van de referentieniveaus wordt duidelijk op welk niveau doelen zouden moeten worden gerealiseerd. Doordat de doelen voor taal en rekenen in de verschillende onderwijssectoren worden afgestemd op het referentiekader, ontstaat meer samenhang tussen de doelen en de eisen die in de verschillende sectoren worden gehanteerd. Het referentiekader is een gemeenschappelijke 'taal' voor het taal- en rekenonderwijs in de opeenvolgende onderwijssectoren. Ook biedt het een gemeenschappelijke meetlat voor het niveau van het onderwijs.

Doordat de toetsing van de referentieniveaus in het voortgezet onderwijs en het middelbaar beroepsonderwijs onderdeel wordt van het examen, gaat het in die sectoren om meer dan alleen maar een peiling van vaardigheden. De uitslag van de peiling heeft gevolgen voor de examenuitslag. In het primair onderwijs heeft de uitslag van de peiling niet zulke consequenties.

Het referentiekader is van kracht sinds schooljaar 2010/11 en geeft dus nu al sturing aan de inrichting van het taal- en rekenonderwijs. Een volgende stap is dat leerlingen zullen worden getoetst op hun beheersing van de referentieniveaus. In het basisonderwijs wordt de verplichte eindtoets die met ingang van 2013 wordt afgenomen stapsgewijs aangepast aan de referentieniveaus, zodat de eindtoets in 2015 een beeld geeft van de beheersing van de referentieniveaus. In het voortgezet onderwijs worden de referentieniveaus voor taal vanaf 2013/14 ingebouwd in het eindexamen Nederlands en wordt voor rekenen een aparte eindtoets afgenomen die meetelt voor het eindexamen. De rekentoets wordt onderdeel van de 'kernvakkenregel' voor het bepalen van de eindexamenuitslag (zie het ontwerp Aanscherping exameneisen voortgezet onderwijs). Aan het einde van onderbouw van het voortgezet onderwijs zal een diagnostische toets afgenomen worden waarmee kan worden nagegaan of leerlingen op de goede weg zitten.

In het mbo zullen de nieuw in te voeren eindexamens taal en rekenen gebaseerd zijn op de referentieniveaus (mbo niveau 4 m.i.v. 2013/14; mbo niveau 2 en 3 m.i.v. 2014/15). Aan de hand van de eerste ervaringen en de uitslagen van eerste peilingen zullen de referentieniveaus de komende jaren nader worden ingevuld.

Scholen wordt op uiteenlopende manieren steun geboden bij het afstemmen van hun onderwijs op de referentieniveaus (deskundigheidbevordering door nascholing, ondersteuning door het Steunpunt taal en rekenen en ondersteuning door educatieve organisaties).

14.3 Doel

De invoering van referentieniveaus heeft als centrale doelstelling het versterken van de taal- en rekenvaardigheden van leerlingen en studenten en het verbeteren van de aansluiting van het onderwijs in deze leergebieden tussen de verschillende sectoren (TK 2009/2010). Meer in het bijzonder gaat het om:

- het duidelijk maken wat van scholen wordt verwacht op het gebied van taal en rekenen;
- het vergroten van de zichtbaarheid van het prestatieniveau;
- het vergroten van de doelgerichtheid van het taal- en rekenonderwijs;
- het bevorderen dat scholen opbrengstgericht werken;
- het realiseren van meer eenduidigheid in het taal- en rekenonderwijs in opeenvolgende onderwijssectoren en het realiseren van een doorlopende leerlijn;
- het verbeteren van de aansluiting en de overdracht van leerlingen tussen opeenvolgende onderwijssectoren.

14.4 Veronderstellingen over gedragsreacties en effecten

De centrale veronderstelling achter het beleid is dat een heldere omschrijving van het gewenste taal- en rekenniveau van leerlingen in de verschillende onderwijssectoren helpt om de taal- en rekenprestaties te verbeteren.

Effecten op instellingsniveau

De veronderstelling is dat de invoering van referentieniveaus veranderingen teweegbrengt in het onderwijs dat scholen aanbieden. Scholen zullen hun taal- en rekenonderwijs aanpassen aan de referentieniveaus, en taal- en rekenonderwijs zal een meer centrale plek krijgen in het onderwijsprogramma. Dat geldt met name voor het mbo, waar taal- en rekenonderwijs van oudsher niet dezelfde prominente plek in het programma heeft als in het basis- en voortgezet onderwijs.

De verwachting is dat de streefniveaus scholen zullen aanzetten tot een meer doelgerichte en opbrengstgerichte aanpak. De grotere aandacht voor met name taal- en rekenonderwijs kan voor scholen aanleiding zijn om ook bij andere onderdelen uit het lesprogramma meer aandacht te gaan besteden aan taal ('schoolbreed taalbeleid'). De introductie van de referentieniveaus zal scholen ertoe aanzetten om hun leerlingen tussentijds te toetsen op hun voortgang op het gebied van taal en rekenen – voor zover ze dat nog niet doen. De eenduidigheid en de transparantie van behaalde resultaten kunnen bevorderen dat scholen zich meer met elkaar gaan vergelijken en op basis daarvan hun doelen bijstellen. Scholen kunnen de scores die zij halen op de referentietoets ook hanteren als marketinginstrument in de concurrentie met andere scholen.

Scholen in het voortgezet onderwijs en het mbo zullen dankzij het referentiekader beter inzicht hebben in de vaardigheden waarmee leerlingen uit het voorafgaande onderwijs instromen. Het afnemen van instroomtoetsen om het niveau van leerlingen vast te stellen is niet meer of minder nodig. Door het onderwijs af te stemmen op het niveau waarop leerlingen instromen, kan leerlingen een doorlopende leerlijn in taal en rekenen worden geboden.

Effecten op leerkrachtniveau

Leraren zullen hun taal- en rekenonderwijs mogelijk moeten aanpassen aan de referentieniveaus. Voor zover dat nog niet gebeurt, zullen zij vaker diagnostische toetsen moeten afnemen om vast te stellen hoe leerlingen ervoor staan. Op scholen waar wordt gekozen voor een 'schoolbrede' invoering van de referentieniveaus zullen ook leraren die andere vakken geven dan Nederlands meer aandacht moeten gaan besteden aan taal. Een deel van de leraren zal behoefte hebben aan bijscholing om bekend te worden met de referentieniveaus en om ermee te leren werken.

Effecten op leerlingenniveau

Leerlingen (en hun ouders) krijgen met de referentieniveaus inzicht in hoe ze ervoor staan op het gebied van taal- en rekenen. Als scholen hun taal- en rekenbeleid aanscherpen en met het oog op de referentieniveaus meer werk maken van opbrengstgericht

werken, zullen de taal- en de rekenvaardigheid van leerlingen naar verwachting verbeteren, met name ook van leerlingen die niet aan de fundamentele norm voldoen. Het ideaal is dat tussen opeenvolgende onderwijssectoren doorlopende leerlijnen worden gerealiseerd, waardoor de aansluiting tussen opeenvolgende onderwijsfasen verbetert. Leerlingen zullen dan vaker op een passend niveau kunnen worden geplaatst en onderwijs krijgen dat aansluit op het niveau waarmee ze het voorafgaande onderwijs hebben afgesloten. Ook bij wisseling van school, bijvoorbeeld vanwege verhuizing, zullen breuken makkelijker kunnen worden vermeden.

In het voortgezet onderwijs en het middelbaar beroepsonderwijs krijgen leerlingen te maken met nieuwe eisen die doorwerken in de examens (rekentoets voortgezet onderwijs, eindexamen taal- en rekenen in het mbo). Voor een deel van de leerlingen en studenten zal dat een verzwaring van het examen betekenen. Als aan deze verzwaring van eisen geen verbetering van de vaardigheden van leerlingen en studenten voorafgaat, zullen hun slagingskansen waarschijnlijk verslechteren. Als scholen naar aanleiding daarvan voorzichtiger worden met de toelating van leerlingen, zullen leerlingen in de toekomst minder makkelijk worden toegelaten tot hogere niveaus van voortgezet onderwijs en middelbaar beroepsonderwijs.

Onbedoelde effecten en beleidsafwijkingen

De invoering van de referentieniveaus kan onbedoelde bijwerkingen hebben. Het centraal stellen van de doelstelling dat zo veel mogelijk leerlingen een bepaald taal-rekenniveau halen, kan tot gevolg hebben dat andere doelen en onderdelen in het lesprogramma minder aandacht gaan krijgen. De nadruk op basisvaardigheden kan leiden tot een versmalling van het aanbod.

Als scholen zich bij de invoering van de referentieniveau vooral richten op het fundamentele referentieniveau en niet op het hogere streefniveau, zou neerwaartse nivellering kunnen optreden. Scholen realiseren dan het fundamentele niveau, maar ook niet meer dan dat. In de praktijk moet blijken of ook het hogere streefniveau dat uitdaging moet bieden aan de betere leerlingen een rol zal gaan spelen in het onderwijs.

Het referentiekader gaat uit van een doorlopende leerlijn, waarbij het programma in de volgende onderwijssector aansluit op waar de leerling in de voorgaande sector is gebleven. Omdat het referentiekader niet vanaf de jongste leeftijd geleidelijk aan wordt ingevoerd, kunnen leerlingen in het voortgezet onderwijs en het middelbaar beroepsonderwijs de komende jaren te maken krijgen met eisen waarop zij in hun voorgaande loopbaan niet zijn voorbereid. Omdat achterstanden in een latere fase (in het voortgezet onderwijs, in het mbo) moeilijk zijn te repareren, kan blijken dat de referentieniveaus te hoog zijn gegrepen in de lagere leerwegen van het vmbo en het mbo. Een te grote kloof tussen het feitelijke en het gewenste niveau kan de legitimiteit van de referentieniveaus aantasten. Aanpassing van de norm aan het feitelijk lagere niveau van leerlingen zou echter afbreuk doen aan de gedachte achter de referentienormen.

14.5 Evaluatie

Effectevaluatie

Om te kunnen vaststellen of de invoering van de referentieniveaus de beoogde effecten heeft, is het essentieel dat een vergelijking kan worden gemaakt tussen scholen of leerlingen die te maken krijgen met de referentieniveaus en een vergelijkbare controlegroep van scholen of leerlingen bij wie dat niet het geval is. Omdat de referentieniveaus integraal worden ingevoerd, is er in principe echter geen controlegroep van scholen of leerlingen die niet met het beleid te maken krijgt. Bij gebrek aan een (vergelijkbare) controlegroep is een robuuste effectevaluatie dus lastig te realiseren.

Een complicerende factor bij de evaluatie van de invoering van de referentienormen is dat de invoering is verweven met andere maatregelen, zoals de verplichte eindtoets en het leerlingvolgsysteem in het primair onderwijs, de invoering van een rekentoets als onderdeel van het eindexamen met aangescherpte slaagregels in het voortgezet onderwijs, de invoering van centrale eindexamens voor taal en rekenen in het mbo en het beleid om het opbrengstgericht werken te bevorderen (prestatiebox po en vo, 'School aan zet'). De verwachtingen over de effecten van de invoering van het referentiekader zijn deels gebaseerd op deze maatregelen, die in samenhang met de referentieniveaus worden genomen. Aan de andere kant bouwen deze maatregelen ook voort op het referentiekader. Zelfs als het lukt om goede vergelijkingsgroepen samen te stellen is het dus lastig veranderingen in prestaties bij taal en rekenen eenduidig toe te schrijven aan de invoering van het referentiekader.

Als beleid integraal wordt ingevoerd, kan de intensiteit van het beleid in de praktijk uiteenlopen voor verschillende betrokkenen. Voor betrokkenen (scholen, leerlingen) die al in hoge mate aan de doelstellingen van een nieuwe maatregel voldoen, heeft die maatregel relatief weinig impact. Voor betrokkenen die nog niet aan de doelstellingen van het beleid voldoen, heeft het beleid meer betekenis. Dergelijke verschillen in intensiteit kunnen worden gebruikt om experimentele en controlegroepen af te bakenen.

Een centrale doelstelling van de invoering van de referentieniveaus is het bevorderen dat leerlingen minimaal over een bepaald niveau van taal- en rekenvaardigheid beschikken. Voor groepen van leerlingen of scholen die bij de invoering van het referentiekader al in ruime mate aan het gevraagde minimum (de referentienorm) voldoen, heeft de referentienorm weinig impact; zij kunnen fungeren als controlegroep. Groepen van leerlingen of scholen die nog weinig aan de referentienorm voldoen, vormen de experimentele groep. Aan de hand van herhaalde prestatiemetingen kan de trend in prestaties in de experimentele en de controlegroep met elkaar worden vergeleken. Een verschil in trend tussen beide groepen is dan een indicatie van het effect van het beleid (difference-in-differencebenadering). Een essentiële aanname hierbij is dat de trend zonder beleid in beide groepen gelijk zou zijn geweest (common-trendsassumptie). Naarmate beide groepen in allerlei opzichten meer van elkaar verschillen, is die aanname problematischer. De veronderstelling bij deze opzet is verder dat het referentiekader vooral aangrijpt aan de onderkant van de prestatieverdeling via de fundamentele norm (F-norm). Als de hogere streefnorm (S-norm) scholen eveneens aanzet tot extra inspanningen op het gebied

van taal en rekenen, is deze benadering niet bruikbaar. De referentienormen hebben dan voor alle scholen en leerlingen impact, en er is dan geen controlegroep van scholen of leerlingen. Verder is deze opzet niet geschikt om het effect van de maatregel vast te stellen op de doelstelling van het realiseren van een doorlopende leerlijn en betere overdracht van leerlingen tussen opeenvolgende onderwijssectoren.

Op welke groepen zou een trendvergelijking zich kunnen richten? In het basisonderwijs is het lastig om goede vergelijkingsgroepen te vinden. Nagegaan zou kunnen worden of prestatieverschillen tussen jongens en meisjes bij taal en rekenen een bruikbaar aanknopingspunt bieden.¹ In het voortgezet onderwijs zijn er mogelijk verschillen tussen leerlingen in de onderscheiden sectoren en profielen in de bovenbouw van vmbo, havo en vwo in de mate waarin ze aan de taal- en rekennormen voldoen. In het mbo zijn er bij rekenen verschillen gevonden in tussen diverse sectoren; in de sector techniek halen meer leerlingen het (voorlopige) referentieniveau dan in de sectoren zorg en welzijn en economie.²

De voorbehouden die bij deze opzet zijn te maken – de sterke verwevenheid met andere maatregelen, de houdbaarheid van de common-trendsassumptie – kunnen bij de verdere uitwerking van het evaluatieontwerp tot de conclusie leiden dat een robuust effectmodel toch niet haalbaar is. De hiervoor geschetste trendvergelijking tussen groepen van scholen en leerlingen krijgt dan het karakter van een monitor. Deze laat zien hoe de prestaties van verschillende groepen leerlingen op verschillende onderwijsniveaus zich ontwikkelen ten opzichte van de referentienormen en ten opzichte van elkaar, maar laat geen harde uitspraken toe over het optreden van effecten.

In het kader van een monitor kan eveneens aandacht worden besteed aan de andere centrale doelstelling van de maatregel, het realiseren van een doorlopende leerlijn voor taal en rekenen (zie verder hieronder bij ‘verklarende evaluatie’).

De taal- en rekenvaardigheid van leerlingen zijn een centrale uitkomstmaat voor de effectevaluatie. Hierbij kan worden aangesloten bij de toetsgegevens van leerlingen die het Cito verzamelt voor de invoering van de referentieniveaus (de eindtoets, diagnostische toetsen). In het voortgezet onderwijs loopt een pilotonderzoek voor de rekentoets die onderdeel wordt van het eindexamen. De gegevens die in 2012/2013 in de pilot worden verzameld, zouden kunnen fungeren als voormeting in de trendanalyse, en de uitkomsten van de ingevoerde rekentoets als nameting. Een voorwaarde is dat de toetsen door de jaren heen vergelijkbaar blijven, of dat in ieder geval een link kan worden gelegd tussen de scores in opeenvolgende jaren. Een probleem kan zijn dat er voor leerlingen die deelnemen aan de pilot nog niets op het spel staat, terwijl het resultaat op de uiteindelijke rekentoets wel consequenties heeft. De uitslag op een ‘low stake’ en een ‘high stake’ test zijn mogelijk niet goed met elkaar te vergelijken.

Verklarende evaluatie

De verklarende evaluatie is bedoeld om de mechanismen te achterhalen die verantwoordelijk zijn voor het optreden (of achterwege blijven) van de beoogde effecten van het

referentiekader. Centraal daarbij staat de vraag hoe de referentieniveaus in de praktijk van het onderwijs zijn ingevoerd. Welke veranderingen in de aanpak van het taal- en rekenonderwijs hebben zich voorgedaan en zijn verbeteringen van prestaties op het gebied van taal en rekenen daaraan toe schrijven? Onderzoek bij scholen en leerkrachten kan hier licht op werpen.

Een eerste vraag is die naar de bekendheid met en de steun voor de referentieniveaus. Zijn scholen en leerkrachten op de hoogte van de referentieniveaus, is er draagvlak voor de uitgangspunten van het referentiekader en voor het zwaarder inzetten op taal- en rekenvaardigheid? Hebben scholen en leerkrachten een beeld van hoe de prestaties van hun leerlingen zich verhouden tot de referentieniveaus? Worden de referentieniveaus haalbaar geacht?

Vervolgens kan worden onderzocht hoe scholen en leerkrachten omgaan met de referentieniveaus. Is het taal- en rekenonderwijs aangepast naar aanleiding van de invoering van de referentieniveaus? Krijgen taal en rekenen meer aandacht, en zo ja, gaat dat ten koste van andere leergebieden? Realiseren scholen en docenten een opbrengstgerichte aanpak en welke rol spelen de referentieniveaus daarbij? Voeren scholen in het voortgezet onderwijs een schoolbreed taal- en rekenbeleid en hangt dat samen met de resultaten die leerlingen halen?

Ook is van belang om na te gaan hoe scholen en leerkrachten omgaan met het fundamentele niveau en het streefniveau. Richt men zich vooral op het halen van het fundamentele niveau met zo veel mogelijk leerlingen of wordt er ook werk gemaakt van het hogere streefniveau?

Bij al deze vragen zullen er verschillen zijn tussen onderwijssectoren en verschillen naar gelang de uitgangssituatie van de school.

Voor het realiseren van de doorlopende leerlijn in opeenvolgende onderwijssectoren is het van belang om te onderzoeken hoe de overdracht van leerlingen verloopt. Stelt het vervolgonderwijs zich op de hoogte van het taal- en rekenniveau waarop leerlingen binnenkomen? Passen ze het lesaanbod daarop aan?

Noten

- 1 Er zijn daarnaast in het basisonderwijs aanzienlijke verschillen in taal- en rekenvaardigheid tussen kinderen uit het hoogopgeleide milieu en kinderen met een niet-westers allochtone achtergrond of kinderen met laagopgeleide ouders. De aanname dat de prestaties van deze groepen een gelijke trend laten zien zonder beleid lijkt echter niet plausibel. Niet-westers allochtone leerlingen lopen hun achterstand geleidelijk aan in, die van autochtone achterstandsleerlingen neemt toe (Gijsberts en Herweijer 2009).
- 2 Zie Cito (2011).

15 Leraren een masteropleiding

15.1 Aanleiding en achtergrond

Het kabinet zet in op versterking van de kwaliteit van het onderwijs. De prestaties van de Nederlandse leerlingen moeten verbeteren, zodat ze zich kunnen meten met de mondiale top. De centrale gedachte van het actieplan ‘Leraar 2020 – een krachtig beroep’ is dat de leraar en de schoolleider cruciaal zijn voor de kwaliteit van het onderwijs, en onmisbaar voor de verbetering ervan. Hierbij wordt verwezen naar bijvoorbeeld de studie van McKinsey (2007) over de rol die leerkrachten spelen in de best presterende onderwijsstelsels. Het verbeteren van de prestaties van leerlingen vereist dus investeringen in de kwaliteit van leerkrachten en schoolleiders. Daarvoor wordt ingezet op een verbetering van de kwaliteit van lerarenopleidingen en op verhoging van de kwaliteit van de zittende leraren.

15.2 Inhoud

In het actieplan ‘Leraar 2020 – een krachtig beroep’ zijn verschillende maatregelen aangekondigd om de professionele kwaliteit van leerkrachten te vergroten. Een daarvan is verhoging van het opleidingsniveau van leraren. In dat kader is overwogen om in de toekomst voor nieuwe leraren een masteropleiding verplicht te stellen, maar vooralsnog wordt daarvan afgezien. Het beleid richt zich op verbetering van de lerarenopleidingen en op verhoging van de kwaliteit van de zittende leraren. Lerarenopleidingen krijgen de mogelijkheid om studenten te selecteren op basis van hun geschiktheid voor het leraarsvak en hun vakkennis.

Voor zittende leraren wordt ervan uitgegaan dat zij hun kwaliteit kunnen verhogen door verdere scholing. De lerarenbeurs en de faciliteiten van het algemene scholingsbeleid bieden daarvoor mogelijkheden. Voldoende geschikt aanbod van opleidingen op masterniveau is hierbij belangrijk. Daarnaast zal worden nagegaan hoe de vraag naar masteropleidingen beter kan worden gearticuleerd, zodat het aanbod van opleidingen beter kan inspelen op de vraag. Onderzoek moet duidelijk maken welke rol masteropgeleide leraren kunnen spelen binnen de school en wat de toegevoegde waarde van is van een masteropleiding voor de kwaliteit van leerkrachten. Het betreft hier leraren in het primair onderwijs, het voortgezet onderwijs en het middelaar beroepsonderwijs. In de bijlage wordt een kort overzicht gegeven van de verschillende varianten van masteropleidingen die van belang kunnen zijn voor leerkrachten.

15.3 Doel

Het doel van het verstrekken van faciliteiten waarmee zittende leraren een masteropleiding kunnen volgen, is het verhogen van de professionele kwaliteit van leraren.

Een hogere professionele kwaliteit van leraren zal leiden tot betere leerresultaten. Dat laatste is te beschouwen als het ultieme, verder weg gelegen doel van de maatregel.

15.4 Veronderstellingen over gedragsreacties en effecten

De centrale veronderstelling is dat de kwaliteit van het onderwijs gebaat is bij een verhoging van het opleidingsniveau van leerkrachten tot masterniveau. Op verschillende niveaus worden hiervan positieve effecten verwacht.

Effecten op leraarniveau

De verwachting is dat de kwaliteit van leraren toeneemt door het volgen van een opleiding op masterniveau. Een belangrijke graadmeter voor de kwaliteit van leraren is hun effectiviteit: het vermogen om goede leerprestaties met leerlingen te halen. Het vermogen om effectief les te geven' en het beheersen van de vakdidactiek en de vakinhoud zijn daarvoor van belang. De verwachting is dat leraren dankzij een masteropleiding in ruimere mate over deze eigenschappen zullen beschikken en daardoor ook betere resultaten met hun leerlingen zullen halen.

Buitenlands onderzoek naar de effectiviteit van leraren – vaak verricht in de vs – laat overigens geen effecten zien van een opleiding op masterniveau (Van Veen et al. 2011). Niet uitgesloten is echter dat een masteropleiding indirect – via bijvoorbeeld het kenmerk 'beheersen van de vakinhoud' – toch invloed heeft op leerprestaties van leerlingen. Het door Van Veen et al. (2011) samengevatte onderzoek geeft onvoldoende uitsluitsel over de relatie tussen een masteropleiding en overige factoren die van invloed zijn op de kwaliteit van lesgeven en de leerprestaties van leerlingen. Onderzoek naar de invloed van de mastergraad op de kwaliteit van leraren in het Nederlandse onderwijs ontbreekt.

Voor leraren zelf kan een masteropleiding ook effecten hebben op hun loopbaan (salaris, functie, mobiliteit). Leraren met een masteropleiding zullen mogelijk aanspraak kunnen maken op een hoger salaris, bijvoorbeeld in het kader van de functiemix, zullen in een aantal gevallen een bredere rol in de school gaan vervullen (bv. in ontwikkeling en vernieuwing van het onderwijs, coördinatie, management) en zullen waarschijnlijk makkelijker van werkgever kunnen veranderen omdat ze gewild zijn op de arbeidsmarkt.

Effecten op schoolniveau

Ook op schoolniveau zijn er mogelijk positieve effecten van masteropgeleide leraren. De aanwezigheid van (meer) masteropgeleide leraren leidt mogelijk tot veranderingen op het organisatorische vlak, in de vorm van meer taak- en functiedifferentiatie. Ook zou een ander onderwijskundig klimaat in de school kunnen ontstaan, doordat masteropgeleide leraren beter in staat zijn bij te dragen aan de onderwijskundige ontwikkeling van de school. De aanwezigheid van masteropgeleide leraren zou ook een 'lerende cultuur' in de school kunnen brengen.

Masteropgeleide leraren kunnen wellicht ook de concurrentiepositie van een school versterken ten opzichte van andere scholen. Scholen zouden het beschikken over masteropgeleide leraren kunnen inzetten als een pr-instrument.

Als veel leraren op masterniveau zijn opgeleid kan dat, ten slotte, ook een positieve uitwerking hebben op het imago en de status van het leraarsberoep. Op de langere termijn zou dit kunnen leiden tot meer wervingskracht en een sterkere concurrentiepositie van het onderwijs op de arbeidsmarkt.

Onbedoelde effecten en beleidsafwijkingen

De centrale gedachte is dat de kwaliteit van leraren verbetert als ze een masteropleiding hebben gevolgd. Doordat er betere leraren voor de klas staan, zullen leerlingen vervolgens beter presteren. Zoals aangegeven, zijn er diverse varianten van masteropleidingen voor leraren met een bacheloropleiding en niet elke combinatie van bachelor- en masteropleiding zal even effectief zijn. Een masteropleiding die slecht aansluit op de gevolgde bacheloropleiding of op de werksituatie van de betrokken leerkracht zal vermoedelijk weinig positief effect sorteren. Of er zich positieve effecten voordoen, zal ook afhangen van de manier waarop leraren met een masteropleiding worden ingezet. Als zij doormoeten naar coördinerende of managementfuncties, kunnen zij alleen nog maar op indirecte wijze – via organisatie en beleid – bijdragen aan de kwaliteit van het lesgeven. Het zou erop kunnen neerkomen dat best geschoolde leraren niet meer voor klas staan. Voor scholen en besturen zal een groei van het aantal masteropgeleide leraren vermoedelijk een stijging van de personeelskosten met zich brengen. Leraren zullen hun hogere opleidingsniveau willen terugzien in hun beloning.

Het opscholen van leraren tot masterniveau kan leiden tot een opwaardering van het leraarsberoep, maar de keerzijde is dat de status van leraren met alleen een bacheloropleiding mogelijk daalt. Er kan een tweedeling ontstaan tussen master- en bacheloropgeleide leraren in scholen met nu hoofdzakelijk bacheloropgeleide leraren.

15.5 Evaluatie

De overheid stelt faciliteiten beschikbaar voor het volgen van een masteropleiding, maar inzicht in de meerwaarde van een masteropleiding voor de kwaliteit van leraren en de kwaliteit van het onderwijs ontbreekt nog. Met een evaluatieonderzoek kan dat inzicht worden vergroot.

Effectevaluatie

De centrale vragen van de effectevaluatie zijn:

- Wat is de bijdrage van een masteropleiding aan de kwaliteit van leraren, niet alleen in termen van relevante kennis en vaardigheden, maar ook in termen van de prestaties die met de leerlingen worden gehaald?
- Wat is het effect van een masteropleiding op de functie en loopbaan van leraren?
- Wat is het effect van meer masteropgeleiden op de schoolorganisatie en de schoolcultuur?

Bij het uitwerken van de effectevaluatie moet er rekening mee worden gehouden dat niet elke type masteropleiding hetzelfde effect zal hebben, dat het effect van een masteropleiding kan variëren naar gelang het soort (bachelor)vooropleiding dat een leraar heeft gevolgd en dat het effect kan verschillen naar gelang de schoolsoort waar de leraar werkzaam is en de functie die de leraar vervult. De vraag naar het effect van een masteropleiding valt dus uiteen in een aantal subvragen:

- Verschilt het effect naar gelang het soort masteropleiding?
- Verschilt het effect naar gelang de onderwijssector waar de leerkracht werkzaam is en de functie die de leerkracht vervult?
- Verschilt het effect naar gelang de vooropleiding van de leerkracht?

Deze drie subvragen moeten in samenhang worden gezien; het gaat om het vaststellen van de effectiviteit van combinaties van bachelor- en masteropleidingen binnen de verschillende onderwijssectoren. In de praktijk zal het aantal combinaties van bachelor- en masteropleidingen binnen de afzonderlijke onderwijssectoren vermoedelijk niet zo groot zijn. Leerkrachten in het basisonderwijs volgen bijvoorbeeld vaak de hbo-master 'special education needs'.

Verder is het van belang om te onderzoeken wat de beste timing is. Is het optimaal om een masteropleiding zo snel mogelijk aansluitend op de bachelor te volgen of is het beter om eerst een aantal jaren praktijkervaring op te doen? Wat de beste timing is, kan verschillen naar gelang het type masteropleiding.

In de beleidstheorie worden zowel effecten op het niveau van leraren en hun leerlingen als op schoolniveau voorspeld. Om vast te stellen of de verwachte effecten zich voordoen, moet een vergelijking worden gemaakt tussen een experimentele groep die de interventie (opsholing naar masterniveau) ondergaat en een controlegroep die de interventie niet ondergaat.

Het is in dat verband essentieel dat de interventie- en de controlegroep goed vergelijkbaar zijn. Als de verdeling over de interventie- en de controlegroep afhangt van factoren die tevens van invloed zijn op de uitkomsten die met de interventie worden beoogd, kunnen geen harde uitspraken worden gedaan over het effect van de interventie.²

Verder moeten we er rekening mee houden dat er in de beleidstheorie zowel effecten worden verwacht op het niveau van leraren en hun leerlingen als effecten op schoolniveau. Er zou dus niet alleen een vergelijking moeten worden gemaakt tussen leraren die een masteropleiding hebben gevolgd en leraren die dat niet hebben gedaan, maar ook tussen scholen met veel en scholen met weinig of geen masteropgeleide leraren. Dat betekent dat zowel op het niveau van leraren als dat van scholen vergelijkingsgroepen moeten worden samengesteld. We verkennen twee mogelijkheden: een aanpak die uitgaat van vergelijkingsgroepen van leraren en een die uitgaat van een experiment waarbij scholen kunnen worden vergeleken.

Vergelijkingsgroepen op het niveau van leraren

Om het effect van een masteropleiding op het niveau van leraren vast te stellen kunnen leraren die een masteropleiding volgen, worden vergeleken met een controlegroep

van leraren die deze opleiding niet volgen. Belangrijk is dat de controlegroep zo veel mogelijk lijkt op de experimentele groep van deelnemers. Idealiter zou de controlegroep moeten bestaan uit leerkrachten die gemotiveerd zijn om een opleiding te volgen maar daar niet aan toekomen. Wachtlijsten of lotingen voor schaarse opleidingsplekken kunnen hiervoor in theorie een aanknopingspunt bieden, maar in de praktijk is dit geen oplossing.³ Wel zou kunnen worden geprobeerd om een controlegroep samen te stellen van leraren die desgevraagd te kennen geven de intentie hebben een masteropleiding te volgen maar zich om praktische redenen niet aanmelden. Deze groep zou kunnen worden opgespoord in een onderzoek onder leerkrachten en daarna als controlegroep kunnen worden ingezet.

Het effectonderzoek kan dan uitgaan van een trendvergelijking op effectmaten tussen beide groepen leraren (effectmaten, zie hieronder). Onder aanname dat de trend in beide groepen zonder interventie gelijk zal zijn, is het verschil in trend toe te schrijven aan de interventie, dat wil zeggen: het volgen en afronden van een masteropleiding (de difference-in-differencebenadering).

Als op deze manier een controlegroep wordt samengesteld, moet worden aangenomen dat 'zeg-gedrag' ('ik zou wel een masteropleiding willen volgen, maar kom er om uitlopende redenen niet toe') een even goede indicatie is van motivatie als het zich daadwerkelijk aanmelden bij een masteropleiding. Op die aanname valt wel wat af te dingen: het antwoord op een vraag naar de intentie om een opleiding te volgen kan worden vertekend door de tendens om sociaal wenselijke antwoorden te geven. Door de controlegroep aanvullend ook op andere relevante kenmerken te matchen met de interventiegroep kan de vergelijkbaarheid tussen controlegroep en interventiegroep worden vergroot, maar het bezwaar van mogelijke vertekening door selectieve samenstelling van de experimentele en controlegroep kan niet helemaal worden ondervangen.

Een experiment op schoolniveau

Met de hiervoor geschetste opzet zijn effecten op schoolniveau moeilijk aan te tonen; de vergelijking gaat uit van groepen leraren, niet van groepen van scholen. Om effecten op schoolniveau te onderzoeken kan een variant van een experimentele opzet worden overwogen. Daarbij zou een groep van scholen kunnen worden benaderd met een aanbod waarbij ze kans maken op extra faciliteiten om hun leraren masteropleidingen te laten volgen. Door slechts een random deel van de scholen die op het aanbod ingaan daadwerkelijk te faciliteren kunnen een experimentele en een controlegroep van scholen worden samengesteld. De scholen die worden uitgeloot, fungeren daarbij als controlegroep.

Om kans te maken op de extra faciliteiten is het een voorwaarde dat scholen die worden uitgeloot zich beschikbaar stellen als lid van de controlegroep.

Belangrijk is dat het totaal aantal scholen dat aan het experiment deelneemt zo groot is dat door random toewijzing van scholen aan de experimentele en de controlegroep daadwerkelijk vergelijkbare groepen van scholen ontstaan. Door scholen te benaderen die op relevante kenmerken niet veel van elkaar verschillen, kan het risico op een niet goed vergelijkbare experimentele en de controlegroep worden verminderd. Het handhaven van het element van loting bij toewijzing aan de beide groepen is belangrijk; als scholen op voorhand geen kans maken op de extra scholingsfaciliteiten zijn zij

waarschijnlijk moeilijk te bewegen om deel uit te maken van de controlegroep, tenzij daar andere faciliteiten tegenover staan.

Ook in de controlegroep zullen er leraren zijn die een masteropleiding volgen, bijvoorbeeld met gebruikmaking van de lerarenbeurs. Het is daarom belangrijk dat de faciliteiten die aan de experimentele groep worden verstrekt een betekenisvolle aanvulling zijn op de reeds beschikbare, en dus leiden tot meer deelname aan masteropleidingen dan in de controlegroep. De vergelijking kan dan uitgaan van verschillen in deelname aan masteropleidingen en/of het aantal masteropgeleide leraren tussen de experimentele en de controlegroep van scholen. Uit voor- en nametingen op effectmaten kan blijken of de trend in beide scholengroepen afwijkt. Alvorens een experiment te starten, zou in een vooronderzoek moeten worden nagegaan of voldoende extra deelname aan masteropleidingen in een experimentele groep kan worden uitgelokt.

De *effectmaten* moeten aansluiten op de doelstellingen en verwachtingen over de opbrengsten van de maatregel. Hierbij is er een onderscheid tussen opbrengsten op het niveau van leraren, op het schoolniveau en op het niveau van leerlingen.

De effecten op het *niveau van leraren* hebben betrekking op:

- de kennis en vaardigheden van leraren van belang voor het realiseren van goede leeropbrengsten bij hun leerlingen;
- de functie, de rol die ze binnen de school vervullen en hun loopbaan (salaris, interne en externe mobiliteit).

Voor de kennis en vaardigheden kan worden aangesloten bij kenmerken die uit de literatuur naar voren komen (vergelijk de kenmerken die Van Veen et al. 2011 bespreken). Op basis van een voor- en een nameting, bijvoorbeeld in de vorm van observaties of beoordeling door leidinggevenden, kunnen veranderingen in kwaliteiten van leraren worden vastgesteld. De keuze van effectmaten moet tevens aansluiten bij het karakter van de gevolgde masteropleiding (onderwijskundig, vakgericht, gericht op management). Veranderingen in de functie en rol van leerkrachten binnen de school kunnen worden vastgesteld door middel van peilingen onder leraren of schoolleiders. Effecten in termen van salaris en mobiliteit op de arbeidsmarkt zijn mogelijk uit registraties af te leiden.

Effecten op *schoolniveau* betreffen:

- het ontstaan taak- en functiedifferentiatie;
- veranderingen in onderwijskundige aanpak onder invloed van de inbreng van masteropgeleide leraren.

In een opzet waarin groepen van scholen worden vergeleken, kunnen veranderingen op schoolniveau worden gepeild bij schoolleiders, zo mogelijk met een voor- en een nameting.

Bij effecten op *leerlingniveau* gaat het om hun leerprestaties. Om effecten vast te stellen moet een maatstaf voor leerwinst worden gehanteerd. In de opzet die uitgaat van de vergelijking van groepen leraren moet een link worden gelegd tussen leraren en hun leerlingen. Hoeveel leerwinst realiseert een leraar met zijn leerlingen, en neemt die leerwinst toe nadat de leraar een masteropleiding heeft gevolgd? Om dat vast te stellen moeten voor- en nametingen van leerwinst worden uitgevoerd bij een experimentele en een controlegroep van leraren. Hierbij mag waarschijnlijk niet worden verwacht dat de effecten in termen van leerwinst zich onmiddellijk na afronden van de masteropleiding voordoen. De keuze van de periode tussen de voor- en de nameting vraagt nadere afweging. Als gebruik kan worden gemaakt van bestaande gegevens uit leerlingvolgsystemen, kunnen verschillende intervallen tussen voor- en nameting worden beproefd. In het voortgezet onderwijs en het mbo hebben leerlingen te maken met verschillende leerkrachten voor verschillende vakken. De vergelijking van leerprestaties moet zich daar richten op een bepaald vak en uitgaan van de toegevoegde waarde die leraren in dat vak bij hun leerlingen realiseren. Te overwegen is om hiervoor aansluiting te zoeken bij de toetsgegevens die worden verzameld in het kader van de referentieniveaus voor taal en rekenen/wiskunde (de effectevaluatie beperkt zich dan tot deze domeinen). Ook leerlingvolgsystemen kunnen een aanknopingspunt bieden om leerwinst te inventariseren.

In de opzet die uitgaat van vergelijking van een experimentele en een controle groep van scholen kunnen de effecten in termen van leerprestaties op schoolniveau worden vastgesteld. Ook dan moet worden uitgegaan van een maatstaf voor leerwinst.

De effectevaluatie richt zich op effecten op verschillende niveaus: bij leraren, in scholen en bij leerlingen. De veronderstelde effecten zullen zich niet allemaal onmiddellijk en tegelijkertijd voordoen. Sommige effecten zullen al op korte termijn kunnen worden waargenomen, andere pas op wat langere termijn. De verdere uitwerking van de effectevaluatie moet daarop worden afgestemd.

Verklarende evaluatie

De beleidstheorie voorspelt effecten op het niveau van leraren, op schoolniveau en in termen van leerprestaties. De verklarende evaluatie moet verhelderen waarom effecten zich al dan niet voordoen. De aandacht gaat daarbij uit naar de manier waarop de maatregel wordt uitgevoerd en de mechanismen die de maatregel op gang brengt. Het onderscheid tussen effect- en verklarende evaluatie is hierbij niet scherp. De effecten van een masteropleiding die op het niveau van leraren zijn te verwachten (meer kennis en vaardigheden) zijn volgens de beleidstheorie de verklaring voor het optreden van effecten op het niveau leerlingen (betere leerprestaties). In de effectevaluatie kunnen effecten op het niveau van leraren en op het niveau van leerlingen met elkaar in verband worden gebracht om de veronderstelde oorzakelijke relatie te toetsen.

In verklarende evaluatie kan de implementatie van de maatregel breder in kaart worden gebracht. Een eerste aandachtspunt is de deelname aan een masteropleidingen door leerkrachten. Als uit de effectevaluatie zou blijken dat een masteropleiding een

toegevoegde waarde heeft, is het belangrijk inzicht te verkrijgen in de factoren die de deelname aan deze opleidingen bevorderen of die juist belemmerend werken. Welke rol spelen persoonlijke omstandigheden, wat is de invloed van de beschikbare faciliteiten en wat is de (ervaren) invloed van het beleid dat schoolleiders en besturen voeren op dit punt? Welk type masteropleiding kiezen leraren, welke overwegingen hebben zij daarvoor en sluit de gekozen opleiding aan op hun huidige functie of is deze juist gericht op mobiliteit naar een andere functie? Een peiling onder leerkrachten, deelnemers en afgestudeerden van masteropleidingen kan licht werpen op deze vragen.

Een tweede aandachtspunt is het beleid van scholen en besturen ten aanzien van masteropgeleide leraren. Is er een beleid om te bevorderen dat leraren een masteropleiding volgen en wat zijn de doelstellingen van dat beleid? Is er beleid voor de inzet van masteropgeleide leraren; is het streven dat ze hun lesgevende taken behouden of worden ze ingezet voor specifiek taken of functies?

Een derde aandachtspunt zijn de ervaringen van masteropgeleide docenten. In de effectevaluatie wordt nagegaan of er veranderingen optreden in de functie en de rol van leerkrachten die een masteropleiding hebben gevolgd. In de verklarende evaluatie kan worden onderzocht hoe leerkrachten hier zelf tegenaan kijken. Hebben ze het gevoel dat een masteropleiding een meerwaarde heeft, en dat ze die meerwaarde kunnen benutten in de klas en in de school?

De mogelijke effecten van meer masteropgeleide leerkrachten op de status en aantrekkingskracht van het leraarsberoep en op de toestroom naar de lerarenopleidingen zijn lastig rechtstreeks vast te stellen. Peilingen onder verschillende groepen (studenten van de lerarenopleidingen, leerkrachten) kunnen wel indicaties geven. Uitgezocht zou kunnen worden of zij een school met veel masteropgeleide leerkrachten als een aantrekkelijke werkomgeving zien.

Bij scholieren die staan voor de keuze van een opleiding in het hoger onderwijs kan worden onderzocht of de mogelijkheid om als leraar een opleiding op masterniveau te volgen aantrekkingskracht uitoefent. Denk bijvoorbeeld aan de aantrekkingskracht van de pabo op vwo-leerlingen als een opleiding op masterniveau een logisch vervolg wordt op de pabo.

Bijlage bij hoofdstuk 15: varianten van masteropleidingen van belang voor leraren

Zowel in het hbo als in het wetenschappelijk onderwijs worden voor (aanstaande) leraren vakinhoudelijke en onderwijskundige masters aangeboden. *Vakinhoudelijke masters* in het hbo richten zich op tweedegraads bevoegde leraren in het voortgezet onderwijs die een eerstegraads bevoegdheid willen halen in het vak dat ze geven (Frans, wiskunde). Voorbeelden van *onderwijskundige masteropleidingen* in het hbo zijn 'special education needs' en 'leren en innoveren'. Zowel docenten in het primair als het voortgezet en het middelbaar beroepsonderwijs kunnen terecht op een onderwijskundige masteropleiding in het hbo.

Op de eerstegraadslerarenopleidingen in het wetenschappelijk onderwijs wordt een *vakinhoudelijke master* gecombineerd met een *onderwijskundige master*. Ook is er de variant waarbij de vakinhoudelijke en de onderwijskundige componenten aansluitend op de bachelor fase gelijk oplopen. Zittende leraren met een tweedegraadsbevoegdheid die bij een universitaire lerarenopleiding een eerstegraads bevoegdheid willen halen, moeten daar zowel een vakinhoudelijke master als een onderwijskundig programma volgen. Verder zijn er universitaire masteropleidingen op het gebied van bijvoorbeeld onderwijskunde en (ortho)pedagogiek. Deze opleidingen geven geen lesbevoegdheid, maar zijn wel relevant voor leraren. Hbo-opgeleide leraren moeten in regel een pre-masterschakelprogramma volgen voor zij worden toegelaten tot een universitair masterprogramma.

Noten

- 1 Het gaat daarbij om kenmerken zoals het beheersen van strategieën om leerlingen tot 'leren leren' te brengen, het hebben van hoge verwachtingen over leerlingen, een cognitief uitdagende aanpak, een ordelijke en positief-ondersteunende lesomgeving, een heldere en gestructureerde manier van lesgeven en een activerende lesaanpak (Van Veen et al. 2011: 9).
- 2 Als bijvoorbeeld de motivatie en inzet van leraren van invloed zijn op zowel de resultaten die ze met hun leerlingen halen als op hun deelname aan een masteropleiding, kunnen de betere resultaten die masteropgeleide leraren met hun leerlingen halen niet zonder meer worden toegeschreven aan een masteropleiding.
- 3 Voor zover bekend zijn er geen wachtlijsten of uitsluitingen vanwege schaarse opleidingsplekken of vanwege budgettaire beperkingen. Zo wordt het overgrote deel van de aanvragen voor de lerarenbeurs toegewezen, terwijl leraren die naast het net van de lerarenbeurs vissen vaak toch een opleiding gaan volgen met steun uit andere bron.

DEEL III

Samenvatting en conclusie

1 Samenvatting van de evaluatieontwerpen

In dit rapport hebben we de mogelijkheden onderzocht om van vijftien verschillende onderwijsbeleidsmaatregelen het beleidseffect te kunnen vaststellen. Verschillende varianten van een effectevaluatie passeerden de revue. Onderscheidend voor de verschillende varianten was de wijze waarop een experimentele en controlegroep worden samengesteld. We hebben onderscheid gemaakt tussen:

- een gecontroleerd random experiment;
- quasi-experimentele opzetten met *vergelijkbare* experimentele en controlegroepen, waarbij gebruik wordt gemaakt van toevallige variaties in de praktijk, faseverschillen in de invoering of vergelijking van grensgevallen;
- quasi-experimentele opzetten met *niet vergelijkbare* experimentele en controlegroepen, waarbij door trendvergelijking of door statistische correctie of matching rekening wordt gehouden met verschillen tussen de experimentele en de controlegroep. Hieronder valt ook de variant waarbij gebruik wordt gemaakt van variaties in beleidsintensiteit;
- vergelijking van de complete populatie zonder een controlegroep.

Omdat bij de laatste variant geen controlegroep beschikbaar is, spreken we hier niet meer van effectevaluatie maar van monitoren.

Tevens hebben we bij elk van de vijftien maatregelen suggesties gedaan voor een aanvullende verklarende evaluatie om de mechanismen op te sporen die het effect teweegbrengen. Ook zonder harde effectmeting blijft de (beleids)theorie het aanknopingspunt voor onderzoek. Voor opgetreden veranderingen – of het uitblijven daarvan – na invoering van de maatregel wordt getracht een verklaring te vinden in theorieën op het gebied van onderwijs.

In het onderstaande overzicht is per maatregel weergegeven van welke variant van effectevaluatie het ontwerp uitgaat. Bij sommige maatregelen zijn meerdere varianten van effectevaluatie geschetst. Bij het verder uitwerken van de ontwerpen kan daar een keuze uit worden gemaakt. Ook kunnen verschillende varianten worden gecombineerd.

Bij vijf van de vijftien maatregelen is geen effectmeting mogelijk. Monitoren van doelstellingen is bij deze maatregelen het hoogst haalbare. Dit geldt voor de prestatiebox voor het primair onderwijs, de prestatiebox voor het voortgezet onderwijs, de prestatiebox voor het mbo, de versterking van de kennisbasis in het hbo middels landelijke toetsen en kwaliteit/prestatieafspraken en profiel in de bekostiging van het hoger onderwijs. Bij de overige tien maatregelen zijn een of meerdere varianten van effectevaluatie uitgewerkt. Bij enkele maatregelen zijn voorstellen gedaan voor het opzetten van een experiment. Ook als effectevaluatie mogelijk is, blijft monitoren zinvol (zie de kolom ‘monitor’). Bij alle vijftien maatregelen geeft het ontwerp een aanzet voor een verklarende evaluatie.

	inzicht maar geen effect		minst harde effectmeting		meest robuuste/harde effectmeting			
	verklarende evaluatie	monitor (cohort-vergelijking)	quasi-experimenteel matchen van niet-equivalente groepen	quasi-experimenteel verschil in beleidsintensiteit	quasi-experimenteel verschil in fasen	quasi-experimenteel equivalente groepen	random experimenteel	
1	•	•						
2	•	•	•					
3	•	•						
4	•	•	•	•				
5	•	•						
6	•	•						
7	•	•	•	•				
8	•	•		•	•			
9	•	•					•	
10	•	•				•		
11	•	•						
12	•	•	•					
13	•	•	•			•		•*
14	•	•		•				
15	•	•					•	

* Specifieke variant van random experimentele opzet.

2 Kanttekeningen

Nevenstaand schema laat zien dat de vijftien maatregelen zich voor een meer of minder robuuste effectmeting lenen dan wel voor het monitoren van ontwikkelingen na invoering van de maatregel. Bij bovengenoemde opzetten willen we enkele kanttekeningen plaatsen die van belang zijn bij een verdere uitwerking van de evaluatieontwerpen en bij het interpreteren van de bevindingen.

Veel beleidsinterventies zijn met elkaar verweven. Er verandert momenteel veel tegelijk in de verschillende sectoren van het onderwijs. De verwevenheid van een beleidsmaatregel met andere beleidsinterventies is groot. Dit maakt het lastig om een specifieke interventie separaat te onderzoeken. Het maakt het bovendien moeilijk om veranderingen die bijvoorbeeld uit een cohortvergelijking naar voren kunnen komen eenduidig toe te schrijven aan een specifieke maatregel. De gevonden veranderingen kunnen immers ook het gevolg zijn van andere interventies die hebben plaatsgevonden.

Doelrealisatie in een multi-level- en multi-actorcontext. Verschillende beleidsinterventies waarop deze studie zich richt, hebben betrekking op collectieve, maatschappelijke doelen (vaak per sector) waaraan de individuele scholen/instellingen, docenten en leerlingen/studenten een bijdrage moeten leveren. De beleidsimplementatie heeft dus niet een eenzijdig top-downkarakter, waarbij uitsluitend de landelijke overheid bepaalt wat schoolorganisaties moeten uitvoeren. Er wordt op meerdere (bestuurlijke) lagen en door meerdere partijen gestuurd (policentrische sturing). Een voorbeeld daarvan zijn de op sectorniveau afgesloten akkoorden over de te realiseren beleidsambities. Het op elkaar afstemmen van de verschillende activiteiten door actoren die actief zijn op verschillende niveaus (het al eerder genoemde policy of goal alignment) is een belangrijke en noodzakelijke voorwaarde voor succes. Voor een effectief onderwijsbeleid is het noodzakelijk dat het beleid van de verschillende geledingen in en rondom de school – soms ook inclusief het lokale onderwijsbeleid – op elkaar worden afgestemd. Dat is nog niet eenvoudig, omdat er bij de verschillende actoren weliswaar vaak sprake is van gedeelde belangen, maar soms ook van tegenstrijdige belangen. Het is dus relevant bij een verdere uitwerking van de evaluatieontwerpen rekening te houden met zo'n meerlagen- en meeractorencontext.

Beleidsdoel niet altijd eenduidig

Per beleidsmaatregel worden vaak meerdere doelen beoogd, hetgeen de evaluatie van een specifieke maatregel niet eenvoudig maakt. In de evaluatie kan blijken dat slechts één of een deel van de beoogde doelen zijn bereikt.

Daarnaast is het zo dat scholen of besturen bij sommige maatregelen die hier centraal staan ruimte hebben om zelf prioriteiten te stellen. Dit geldt bijvoorbeeld voor de zogenoemde prestatieboxen primair en voortgezet onderwijs, waarin vijf doelstellingen zijn opgenomen waaruit scholen hun eigen lijstje kunnen samenstellen. In de praktijk zal

het niet eenvoudig zijn om het effect van zo'n meervoudige maatregel te onderzoeken. In het onderzoeksdesign moet immers recht worden gedaan aan de keuze die scholen maken binnen een breder palet van mogelijke doelen, en scholen die verschillende keuzen maken zijn lastig onderling te vergelijken.

Meest robuuste effectevaluatie niet altijd mogelijk

Voor een robuuste effectevaluatie moet een vergelijking kunnen worden gemaakt tussen een experimentele groep die de interventie ondergaat en een bij voorkeur gelijkwaardige controlegroep die niet aan de maatregel wordt blootgesteld. Vaak geldt het random gecontroleerde experiment als de 'gouden standaard', maar ook quasi-experimentele designs bieden mogelijkheden voor een robuuste effectevaluatie.

Van de mogelijkheid om op basis van experimenten te onderzoeken of een maatregel het gewenste effect heeft wordt door de beleidsmakers nog weinig gebruikgemaakt. Veel van de besproken maatregelen worden ingevoerd zonder voorafgaande experimenten om te onderzoeken of ze het gewenste effect hebben. Vaak ontbreekt ook de tijd daarvoor. Bij sommige maatregelen zijn er wel pilots geweest of lopen die nog, maar deze pilots zijn niet opgezet met het oog op een echte effectevaluatie. Doelen die met deze pilots zijn beoogd, betreffen eerder het voorbereiden van het onderwijsveld door het te laten wennen aan de maatregel of het zoeken naar een juiste ijking van een instrument.

Een probleem voor de effectevaluatie is de afwezigheid van controlegroepen. Veel maatregelen worden integraal voor een hele sector ingevoerd. Er zijn desondanks mogelijkheden om een experimentele groep en een controlegroep te onderscheiden, al levert dit een minder robuuste effectmeting op als deze niet gelijkwaardig zijn. Om toch te kunnen vergelijken moeten dan bepaalde aannames worden gedaan. De conclusies die kunnen worden getrokken over het effect van een maatregel zijn daardoor minder robuust. De zogenoemde gouden standaard is doorgaans niet haalbaar bij deze evaluatieontwerpen.

Zonder echte controlegroepen moet een evaluatie uitgaan van een vergelijking van cohorten (voor en na het van kracht worden van de maatregel). Daarmee is echter niet gezegd dat eventueel gevonden veranderingen inderdaad zijn toe te schrijven aan de maatregel. Ook andere factoren hebben invloed en bovendien worden soms verschillende maatregelen vrijwel gelijktijdig ingevoerd.

Aan de voorwaarden voor een harde effectmeting kan in de praktijk vaak niet worden voldaan. Maar ook dan zijn er, zoals we in deel II zagen, verschillende andere onderzoeksmogelijkheden.

Beleidsmakers hebben soms een andere informatiebehoefte dan uitsluitend harde effectmetingen

Uiteraard is de overheid erbij gebaat om te weten of getroffen maatregelen het beoogde effect zullen hebben. Dat neemt niet weg dat veel beleidsmakers allereerst geïnteresseerd zijn in informatie over de implementatie van de maatregel in het veld. Doen zich daar problemen voor, dan kan het beleid of het implementatieproces ex durante immers nog worden bijgesteld. Daarom is voor beleidsmakers de procesevaluatie heel belangrijk. Men wil weten hoe de maatregel 'landt' op scholen en instellingen. Leidt het beleid

tot meer bewustwording bij degenen die de maatregel in de praktijk moeten brengen en gaat men in scholen/instellingen aantoonbaar andere dingen doen, verandert de handelwijze van leerkrachten en docenten?

En hoewel het monitoren van ontwikkelingen niet tot het (harde) effectmeting instrumentarium behoort, laten zulke gegevens wel zien of het uiteindelijk doel van de meeste van deze maatregelen – het op een hoger niveau brengen van de onderwijsprestaties – dichterbij komt.

3 Ten slotte

Het is een goede zaak om te streven naar meer op wetenschappelijk onderzoek gefundeerd beleid. Het valt daarom te prijzen dat beleidsmakers in toenemende mate de eigen beleidstheorieën kritisch tegen het licht houden en de effecten van het beleid empirisch willen onderzoeken. Dat ‘harde bewijzen’ daarbij de voorkeur genieten, lijkt logisch, maar hard bewijs is in evaluatieonderzoek niet altijd te verkrijgen. De praktijk is immers weerbarstig (Slob en Staman 2012) en de processen van data verzamelen en beleid maken lopen doorgaans niet parallel. In het evidence based onderzoek moet er bovendien voor worden gewaakt dat de discussie over causaliteit niet uitsluitend de gekozen methoden en methodologieën betreft. Theoretische verklaringen van waargenomen veranderingen zijn minstens zo belangrijk (Van der Werfhorst 2012). Het is verder ook maar de vraag of een robuuste effectmeting voor beleidsmakers altijd de op dat moment meest nuttige informatie oplevert. Effecten zijn immers pas op langere termijn zichtbaar, terwijl er op kortere termijn behoefte is aan inzicht in veranderingen in het gedrag van betrokken actoren in de school en klas. Wij willen dan ook benadrukken dat het streven naar hard bewijs geen keurslijf moet worden, waardoor voor andere vormen van evaluatieonderzoek geen plaats meer zou zijn.

Er is een stroming in het evaluatieonderzoek die wordt aangeduid met de term ‘realistische evaluatie’. In deze optiek luidt de centrale vraag niet: ‘Wat werkt? of ‘Heeft deze interventie het beoogde effect?’, maar: ‘Wat werkt voor wie, in welke omstandigheden, in welke zin, of op welke wijze?’ (Pawson en Tilly 2004; Pawson 2006). De realistische benadering erkent het belang van harde bewijzen, maar ook van een aanvullende verklarende evaluatie, waarin meer expliciete aandacht is voor de onderliggende mechanismen en voor de (verschillende) context(en) waarin deze mechanismen wel of niet leiden tot het effect (Pater et al. 2012).

In dit rapport hebben we steeds per beleidsinterventie aangeven welke evaluatiemogelijkheden volgens ons mogelijk en zinvol zijn. Soms kon de vraag naar ‘wat werkt’ met een bepaalde onderzoeksopzet beantwoord worden. In andere gevallen lag het accent op het monitoren van ontwikkelingen en het in verschillende contexten systematisch verzamelen van gegevens over wat de beleidsinterventie in het onderwijs teweegbrengt. Deze laatste variant levert misschien geen harde bewijzen op, maar wel bijzonder nuttige informatie over de implementatie en impact van de betreffende interventie.

Literatuur

- Bestuursakkoord primair onderwijs 2012-2015*. Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.
- Borghans, L., F. Cörvers, Th. Dohmen, B. Golsteyn en E. van der Holst (2009). *Vooronderzoek experiment studiekeuzebegeleiding*. Maastricht: Universiteit Maastricht / Netwerk Sociale Innovatie.
- Bosker, R. (2009). Experimentele designs in onderwijsonderzoek. In: ministerie van o c w (red.), *Bewezen beleid in het onderwijs. Essaybundel* (p. 26-39). Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.
- Cito (2011). *Meting taal en rekenen 2010. Een indicatie van leerprestaties in termen van het referentiekader*. Arnhem: Cito.
- Elk, R. van, F.B. van der Meer, M. van der Steeg en D. Webbink (2011). *Zicht op effectiviteit van beleid. Studie naar evaluatieontwerpen voor onderwijs- en wetenschapsmaatregelen*. Den Haag: Centraal Planbureau (CPB Achtergronddocument).
- Gijsberts, M. en L. Herweijer (2009). Onderwijs en opleidingsniveau. In: J. Dagevos en M. Gijsberts (red.), *Jaarrapport integratie 2009* (p. 94-138). Den Haag: Sociaal en Cultureel Planbureau.
- HM Treasury (2011). *The magenta book. Guidance for evaluation*. Geraadpleegd via: www.hm-treasury.gov.uk/d/magenta_book_combined.pdf
- Hofman, R.H., R. Vandenberghen en B.J. Dijkstra (2008). *Kwaliteitszorg, innovatie en schoolontwikkeling. BOPO review. Eindrapport*. Groningen: GION.
- Jansen, F.J.G. en I. de Wolf (2007). Draagt onderwijstoezicht bij aan schoolverbetering? In: *S&B*, jg. 3, nr. 12, oktober.
- Luginbuhl, R. Webbink, D. en I. de Wolf (2007). *Measuring the Effect of School Inspections on Primary School Performance. A Study Based on Cito Test Scores*. Den Haag: Centraal Planbureau (CPB Discussion Paper).
- McKinsey & Company (2007). *How the best-performing school systems com out on top*. Londen: McKinsey & Company.
- Expertgroep doorlopende leerlijnen taal en rekenen (2008). *Over de drempels met taal en rekenen*. Enschede: Expertgroep Doorlopende leerlijnen Taal en Rekenen.
- Onderwijsraad (2006). *Naar meer evidence based onderwijs*. Den Haag: Onderwijsraad.
- o c w (2011a). *Actieplan Basis voor presteren. Naar een ambitieuze leercultuur voor alle leerlingen*. Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.
- o c w (2011b). *Actieplan Beter Presteren: opbrengstgericht en ambitieus. Het beste uit leerlingen halen*. Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.
- o c w (2011c). *Actieplan mbo. Focus op vakmanschap 2011-2015*. Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.
- o c w (2011d). *Kwaliteit in verscheidenheid. Strategische Agenda Hoger onderwijs, Onderzoek en Wetenschap*. Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.
- o c w (2011e). *Leraar 2020 – een krachtig beroep!* Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.
- o c w (2012). *Betref De staat van het hoger beroepsonderwijs*. Brief aan de Tweede Kamer van 17 januari 2012. Referentie 359009.
- o c w, beleidsreactie op eindrapporten alternatieve afstudeertrajecten, 20 mei 2011.
- Pater, C., H. Slighte en E. van Eck m.m.v. M. Derriks (2012). *Verklarende evaluatie. Onderzoek naar de ontwikkeling van een methodiek*. Amsterdam: Kohnstamm Instituut.

- Pawson, R. en N. Tilly (2004). *Realistic evaluation*. Londen: SAGE.
- Pawson, R. (2006). *Evidence-based policy. A realist perspective*. Londen: SAGE.
- Sanders, P.F. en G.J. Visser (2008). *Onderzoek Standaardisering Examinering*. RCEC. Arnhem/Enschede: Cito/ Universiteit Twente.
- Slob, M. en J. Staman (2012). *Beleid en het bewijsbeest. Een verkenning van verwachtingen en praktijken rond evidence based policy*. Den Haag: Rathenau Instituut.
- Torenvlied, R. (1996). *Besluiten in uitvoering: theorieën over beleidsuitvoering modelmatig getoetst op sociale vernieuwing in drie gemeenten* (proefschrift Rijksuniversiteit Groningen). Amsterdam: Thesis.
- TK (2007/2008). *Parlementair onderzoek onderwijsvernieuwingen. 'Tijd voor onderwijs'*. Tweede Kamer, vergaderjaar 2007/2008, 31007, nr. 6.
- TK (2009/2010). *Memorie van toelichting Wet referentieniveaus Nederlandse taal en rekenen*. Tweede Kamer, vergaderjaar 2009/2010, 32290, nr. 3.
- Veen, K. van, J. van Driel en I. Veldman (2011). *Relaties tussen masterniveau en kwaliteit van onderwijs: Conclusies en verklaringen vanuit onderzoek* (paper geschreven in opdracht van het ministerie van o c w, directie Kennis, september 2011). Leiden: ICLO N/Expertisecentrum Leren van Docenten.
- Vossen, R. (vz) (2005). *Een waarneembare verandering in hetgeen met de toezichts- en handhavingsactiviteiten van de inspectie wordt beoogd als gevolg van de activiteiten van de inspectie*. Den Haag: IG-beraad.
- Waslander, S. (2007). *Leren over innoveren*. Utrecht: vo-project Innovatie.
- Webbink, H.D. (2011). *Van evidentie naar impact* (oratie). Rotterdam: Erasmus Universiteit Rotterdam.
- Werfhorst, H. van de (2012). *Tussen beleidsmakers en wetenschapsfilosofie*. In: *Sociologie Magazine*, jg. 20, nr. 2, p. 30-31.
- Yperen, T. van en B. Bijl (2009). *Onderwijs op de ladder: maatwerk in het onderzoek naar effectiviteit*. In: ministerie van o c w (red.), *Bewezen beleid in het onderwijs. Essaybundel* (p. 84-99). Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.

Publicaties van het Sociaal en Cultureel Planbureau

Werkprogramma

Het Sociaal en Cultureel Planbureau stelt twee keer per jaar zijn Werkprogramma vast. De tekst van het lopende programma is te vinden op de website van het SCP: www.scp.nl.

SCP-publicaties

Onderstaande lijst bevat een selectie van publicaties van het Sociaal en Cultureel Planbureau. Deze publicaties zijn verkrijgbaar bij de boekhandel, of via de website van het SCP. Een complete lijst is te vinden op www.scp.nl/publicaties.

Sociaal en Cultureel Rapporten

Investeren in vermogen. Sociaal en Cultureel Rapport 2006. ISBN 90-377-0285-6

Betrekkelijke betrokkenheid. Studies in sociale cohesie. Sociaal en Cultureel Rapport 2008.

ISBN 978 90 377 0368-9

Wisseling van de wacht: generaties in Nederland. Sociaal en Cultureel Rapport 2010. Andries van den Broek, Ria Bronneman-Helmers en Vic Veldheer (red.). ISBN 978 90 377 0505 8

SCP-publicaties 2011

2011-1 *KLEUR. SCP-nieuwjaarsuitgave 2011* (2011). ISBN 978 90 377 0537 9

2011-2 *Stemming onbestemd. Tweede verdiepingstudie Continu Onderzoek Burgerperspectieven* (2011). Paul Dekker en Josje den Ridder (red.). ISBN 978 90 377 0528 7

2011-3 *Vluchtelingengroepen in Nederland. Over de integratie van Afghaanse, Iraakse, Iraanse en Somalische migranten* (2011). Jaco Dagevos en Edith Dourleijn (red.).

ISBN 978 90 377 0526 3

2011-4 *Emancipatiemonitor 2010* (2011). Ans Merens, Marion van den Brakel-Hofmans, Marijke Hartgers en Brigitte Hermans (red.). ISBN 978 90 377 0503 4

2011-5 *Moelijk werken. Gezondheid en de arbeidsdeelname van migrantenvrouwen* (2011). Myra Keizer en Saskia Keuzenkamp. ISBN 978 90 377 0524 9

2011-6 *Informele groepen. Verkenningen van eigentijdse bronnen van sociale cohesie* (2011). E. van den Berg, P. van Houwelingen en J. de Hart (red.). ISBN 978 90 377 0527 0

2011-7 *Gezinsrapport 2011* (2011). Freek Bucx (red.). ISBN 978 90 377 0538 6

2011-8 *Hoe cultureel is de digitale generatie? Het internetgebruik voor culturele doeleinden onder schoolgaande tieners* (2011). Marjon Schols, Marion Duimel en Jos de Haan.

ISBN 978 90 377 0457 0

2011-10 *Kwetsbare ouderen* (2011). Cretien van Campen (red.). ISBN 978 90 377 0542 3

2011-11 *Minder voor het midden. Profijt van de overheid in 2007* (2011). Evert Pommer (red.), Jedid-Jah Jonker, Ab van der Torre, Hetty van Kampen. ISBN 978 90 377 0437 2

2011-12 *Wonen, wijken en interventies. Krachtwijkenbeleid in perspectief* (2011). Karin Wittebrood, Matthieu Permentier, m.m.v. Fenne Pinkster. ISBN 978 90 377 0065 7

2011-13 *Armoedegrens op basis van de budgetbenadering – revisie 2010* (2011). Arjan Soede.

ISBN 978 90 377 0551 5

2011-14 *Werkgevers over de crisis* (2011). Edith Josten. ISBN 978 90 377 0543 0

2011-15 *Op weg naar een inclusieve arbeidsmarkt. Bijdragen van de sprekers op het symposium 15 oktober 2010, Den Haag* (2011). Gerda Jehoel-Gijsbers (red.). ISBN 978 90 377 0546 1

- 2011-16 *Eropuit! Nederlanders in hun vrije tijd buitenshuis* (2011). Desirée Verbeek en Jos de Haan. ISBN 978 90 377 0547 8
- 2011-17 *De opmars van het pgb. De ontwikkeling van het persoonsgebonden budget in nationaal en internationaal perspectief* (2011). K. Sadiraj, D. Oudijk, H. van Kempen, J. Stevens. ISBN 978 90 377 0557 7
- 2011-19 *Kwetsbare ouderen in de praktijk. Een journalistieke samenvatting* (2011). Malou van Hintum. ISBN 978 90 377 0555 3
- 2011-20 *Dimensies van sociale uitsluiting. Naar een verbeterd meetinstrument* (2011). Stella Hoff en Cok Vrooman. ISBN 978 90 377 0532 4
- 2011-21 *Chinese Nederlanders. Van horeca naar hogeschool* (2011). Mérove Gijsberts, Willem Huijnk, Ria Vogels (red.). ISBN 978 90 377 0529 4
- 2011-22 *Gemengd leren. Etnische diversiteit en leerprestaties* (2011). Lex Herweijer. ISBN 978 90 377 0575 1
- 2011-23 *Voorbestemd tot achterstand? Armoede en sociale uitsluiting in de kindertijd en 25 jaar later* (2011). Maurice Guiaux m.m.v. Annette Roest en Jurjen Idema. ISBN 978 90 377 0577 5
- 2011-24 *Kinderen en internetrisico's. EU Kids Online onderzoek onder 9-16-jarige internetgebruikers in Nederland* (2011). Nathalie Sonck en Jos de Haan. ISBN 978 90 377 0576 8
- 2011-25 *De basis meester. Onderwijskwaliteit en basisvaardigheden* (2011). Monique Turkenburg. ISBN 978 90 377 0574 4
- 2011-26 *Acceptatie van homoseksualiteit in Nederland 2011. Internationale vergelijking, ontwikkelingen en actuele situatie* (2011). Saskia Keuzenkamp. ISBN 978 90 377 0579 9
- 2011-27 *Poolse migranten. De positie van Polen die vanaf 2004 in Nederland zijn komen wonen* (2011). Jaco Dagevos (red.). ISBN 978 90 377 0530 0
- 2011-28 *Gewoon aan de slag? De sociale veiligheid van de werkplek voor homoseksuele mannen en vrouwen* (2011). Saskia Keuzenkamp en Ans Oudejans. ISBN 978 90 377 0581 2
- 2011-29 *Acceptance of homosexuality in the Netherlands 2011. International comparison, trends and current situation* (2011). Saskia Keuzenkamp. ISBN 978 90 377 0580 5
- 2011-30 *Nederland in een dag. Tijdsbesteding in Nederland vergeleken met die in vijftien andere Europese landen* (2011). Mariëlle Cloin, Carlijn Kamphuis, Marjon Schols, Annet Tiessen-Raaphorst en Desirée Verbeek. ISBN 978 90 377 0405 1
- 2011-31 *Overheid en onderwijsbestel. Beleidsvorming rond het Nederlandse onderwijsstelsel (1990-2010)* (2011). Ria Bronneman-Helmers. ISBN 978 90 377 0567 6
- 2011-32 *Frail older persons in the Netherlands* (2011). Cretien van Campen (ed.). ISBN 978 90 377 0553 9
- 2011-33 *Maten voor gemeenten. Een analyse van de prestaties van de lokale overheid* (2011). Evert Pommer en Ingrid Ooms, m.m.v. Saskia Jansen, Jedid-Jah Jonker, Hetty van Kempen en Klarita Sadiraj. ISBN 978 90 377 0585 0
- 2011-34 *Maak het nieuw! Over religieuze ontwikkelingen en de positie van de kerken: een persoonlijke geschiedenis* (2011). Joep de Hart. ISBN 978 90 377 0592 8
- 2011-35 *Oudere migranten. Kennis en kennislacunes* (2011, elektronische publicatie). Maaike den Draak en Mirjam de Klerk. ISBN 978 90 377 0597 3
- 2011-36 *Zorg in de laatste jaren. Gezondheid en hulpgebruik in verzorgings- en verpleeghuizen 2000-2008* (2011). Mirjam de Klerk. ISBN 978 90 377 0586 7
- 2011-37 *Jeugdzorg in groeifase. Ontwikkelingen in gebruik en kosten van de jeugdzorg* (2011). Evert Pommer, Hetty van Kempen en Klarita Sadiraj. ISBN 978 90 377 0587 4
- 2011-38 *Verlofvragen. De behoefte aan en het gebruik van verlofregelingen* (2011). Edith de Meester en Saskia Keuzenkamp. ISBN 978 90 377 0589 8

- 2011-39 *De sociale staat van Nederland 2011* (2011). Rob Bijl, Jeroen Boelhouwer, Mariëlle Cloïn, Evert Pommer et al. ISBN 978 90 377 0558 4
- 2011-40 *Kunnen meer kinderen meedoen? Verandering in de maatschappelijke deelname van kinderen, 2008-2010* (2011). Annette Roest. ISBN 978 90 377 0570 6
- 2011-41 *Samenvatting Overheid en onderwijsbestel*. Ria Bronneman-Helmers. ISBN 978 90 377 0590 4

SCP-publicaties 2012

- 2012-1 *Niet alle dagen feest. Nieuwjaarsuitgave 2012* (2012). Paul Schnabel (red.). ISBN 978 90 377 0598 0
- 2012-2 *Waar voor ons belastinggeld? Prijs en kwaliteit van publieke diensten* (2012). Bob Kuhry en Flip de Kam (red.). ISBN 978 90 377 0596 6
- 2012-3 *Jaarrapport integratie 2011* (2012). Mérove Gijsberts, Willem Huijnk en Jaco Dagevos (red.). ISBN 978 90 377 0565 2
- 2012-4 *Bijzondere mantelzorg. Ervaringen van mantelzorgers van mensen met een verstandelijke beperking of psychiatrische problematiek* (2012). Y. Wittenberg, M.H. Kwekkeboom en A.H. de Boer. ISBN 978 90 377 0566 9
- 2012-5 *VeVeRa-IV. Actualisatie en aanpassing ramingsmodel verpleging en verzorging 2009-2030* (2012). Evelien Eggink, Debbie Oudijk en Klarita Sadiraj. ISBN 978 90 377 0594 2
- 2012-6 *Van Anciaux tot Zijlstra. Cultuurparticipatie en cultuurbeleid in Nederland en Vlaanderen* (2012). Quirine van der Hoeven. ISBN 978 90 377 0583 6
- 2012-7 *Vraag naar arbeid 2011* (2012). Edith Josten, Jan Dirk Vlasblom, Marian de Voogd-Hamelink. ISBN 978 90 377 0601 7
- 2012-8 *Measuring and monitoring immigrant integration in Europe* (2012). Rob Bijl en Arjen Verweij (red.) ISBN 978 90 377 0569 0
- 2012-9 *IQ met beperkingen. De mate van versandelijke handicap van zorgvragers in kaart gebracht* (2012). Isolde Woittiez, Michiel Ras en Debbie Oudijk. ISBN 978 90 377 0602 4
- 2012-10 *Niet te ver uit de kast. Ervaringen van homo- en biseksuelen in Nederland* (2012). Saskia Keuzenkamp (red.), Niels Kooiman, Jantine van Lisdonk. ISBN 978 90 377 0603 1
- 2012-11 *The Social State of the Netherlands 2011. Summary* (2012). Rob Bijl, Jeroen Boelhouwer, Mariëlle Cloïn en Evert Pommer (red.) ISBN 978 90 377 0605 5
- 2012-13 *Sturen op geluk. Geluksbevordering door nationale overheden, gemeenten en publieke instellingen* (2012). Cretien van Campen, Ad Bergsma, Jeroen Boelhouwer, Jacqueline Boerefijn, Linda Bolier. ISBN 978 90 377 0608 6
- 2012-14 *Countries compared on public performance. A study of public sector performance in 28 countries* (2012). Jedid-Jah Jonker (red.). ISBN 978 90 377 0584 3
- 2012-15 *Versterking data-infrastructuur sport* (2012). Annet Tiessen-Raaphorst en Jos de Haan. ISBN 978 90 377 0613 0
- 2012-16 *De sociale staat van de gemeente. Lokaal gebruik van de scp-leefsituatie-index* (2012). Jeroen Boelhouwer (SCP), Rob Gilsing (Verwey-Jonker Instituut). ISBN 978 90 377 0612 3
- 2012-17 *Belemmerd aan het werk. Trendrapportage ziekteverzuim, arbeidsongeschiktheid en arbeidsdeelname personen met gezondheidsbeperkingen* (2012). Maroesjka Versantvoort en Patricia van Echtelt (red.). ISBN 978 90 377 0616 1
- 2012-18 *Monitor Talent naar de Top 2011* (2012). Ans Merens (red.) en Commissie Monitoring Talent naar de Top. ISBN 978 90 377 0610 9
- 2012-20 *Verzorgd uit de bijstand. De rol van gedrag, uiterlijk en taal bij de re-integratie van bijstandsvangers* (2012). Patricia van Echtelt en Maurice Guiaux. ISBN 978 90 377 0614 7

- 2012-21 *De virtuele kunstkar. Cultuurdeelname via oude en nieuwe media* (2012). Nathalie Sonck en Jos de Haan. ISBN 978 90 377 0619 2
- 2012-22 *Op zoek naar bewijs. Evaluatieontwerpen onderwijsmaatregelen* (2012). Lex Herweijer en Monique Turkenburg. ISBN 978 90 377 0618 5

SCP-essays

- 1 *Voorbeelden en nabeelden* (2005). Joep de Hart. ISBN 90 377 0248-1
- 2 *De stem des volks* (2006). Arjan van Dixhoorn. ISBN 90 3770265-1
- 3 *De tekentafel neemt de wijk* (2006). Jeanet Kullberg. ISBN 90 377 0261 9
- 4 *Leven zonder drukte* (2006). Tjirk van der Ziel met een naschrift van Anja Steenbekkers en Carola Simon. ISBN 90 377 0262 7
- 5 *Otto Neurath en de maakbaarheid van de betere samenleving* (2007). Ferdinand Mertens. ISBN 978 90 5260 260 8

Overige publicaties

- Hoe het ons verging... Traditionele nieuwjaarsuitgave van het SCP* (2010). Paul Schnabel (red.). ISBN 978 90 377 0465 5
- Wmo Evaluatie. Vierde tussenrapportage. Ondersteuning en participatie van mensen met een lichamelijke beperking; twee jaar na de invoering van de Wmo* (2010). A. Marangos, M. Cardol, M. Dijkgraaf, M. de Klerk. ISBN 978 90 377 0470 9
- Op weg met de Wmo. Journalistieke samenvatting door Karolien Bais. Mirjam de Klerk, Rob Gilsing en Joost Timmermans. Samenvatting door Karolien Bais* (2010). ISBN 978 90 377 0469 3
- NL Kids online. Risico's en kansen van internetgebruik onder jongeren* (2010). Jos de Haan. ISBN 978 90 377 0430 3
- Kortdurende thuiszorg in de Awbz. Een verkenning van omvang, profiel en afbakening* (2010). Maaikeden Draak. ISBN 978 90 377 0471 6
- De publieke opinie over kernenergie* (2010). Paul Dekker, Irene de Goede, Joop van der Pligt. ISBN 978 90 377 0488 4
- Op maat gemaakt? Een evaluatie van enkele responsverbeterende maatregelen onder Nederlanders van niet-westerse afkomst* (2010). Joost Kappelhof. ISBN 978 90 377 0495 2
- Oudere tehuisbewoners. Landelijk overzicht van de leefsituatie van ouderen in instellingen 2008-2009* (2010). Maaikeden Draak. ISBN 978 90 377 0499 0
- Kopers in de knel? Een scenariostudie naar de gevolgen van de crisis voor huiseigenaren met een hypotheek* (2010). Michiel Ras, Ingrid Ooms, Evelien Eggink. ISBN 978 90 377 0498 3
- Gewoon anders. Acceptatie van homoseksualiteit in Nederland* (2010). Saskia Keuzenkamp. ISBN 978 90 377 0502 7
- De aard, de daad en het Woord. Een halve eeuw opinie- en besluitvorming over homoseksualiteit in protestants Nederland, 1959-2009* (2010). David Bos. ISBN 978 90 377 0506 5
- Werkloosheid in goede banen. Bijdragen aan de SCP-studiemiddag 2010* (2010). Patricia van Echtelt (red.). ISBN 978 90 377 0516 4
- Europa's welvaart. De Lissabon Agenda in een breder welvaartspectief en de publieke opinie over de Europese Unie* (2010). Harold Creusen (CPB), Paul Dekker (SCP), Irene de Goede (SCP), Henk Kox (CPB), Peggy Schijns (SCP) en Herman Stolwijk (CPB). ISBN 978 90 377 0492 1
- Maakt de buurt verschil?* (2010). Merové Gijsberts, Miranda Vervoort, Esther Havekes en Jaco Dagevos. ISBN 978 90 377 0227 9

- Mantelzorg uit de doeken* (2010). Debbie Oudijk, Alica de Boer, Isolde Woittiez, Joost Timmermans, Mirjam de Klerk. ISBN 978 90 377 0486 0
- Monitoring acceptance of homosexuality in the Netherlands* (2010). Saskia Keuzenkamp. ISBN 978 90 377 0484 6
- Registers over wijken* (2010). Matthieu Permentier en Karin Wittebrood (SCP), Marjolijn Das en Gelske van Daalen (CBS). ISBN 978 90 377 0499 0
- Data voor scenario's en ramingen van de GGZ* (2010). Cretien van Campen. ISBN 978 90 377 0494 5
- Continu Onderzoek Burgerperspectieven. Kwartaalbericht 2010. Deel 1* (2010). Paul Dekker, Josje den Ridder en Irene de Goede. ISBN 978 90 377 0490 7
- Continu Onderzoek Burgerperspectieven. Kwartaalbericht 2010. Deel 2* (2010). Josje den Ridder en Paul Dekker. ISBN 978 90 377 0507 2
- Continu Onderzoek Burgerperspectieven. Kwartaalbericht 2010. Deel 3* (2010). Josje den Ridder, Lonneke van Noije en Eefje Steenvoorden. ISBN 978 90 377 0508
- Continu Onderzoek Burgerperspectieven. Kwartaalbericht 2010. Deel 4* (2010). Josje den Ridder, Paul Dekker en Eefje Steenvoorden. ISBN 978 90 377 0531 7
- Continu Onderzoek Burgerperspectieven. Kwartaalbericht 2011. Deel 1* (2011). Eefje Steenvoorden, Paul Dekker en Pepijn van Houwelingen. ISBN 978 90 377 0549 2
- Continu Onderzoek Burgerperspectieven. Kwartaalbericht 2011. Deel 2* (2011). Josje de Ridder, Paul Dekker en Pepijn van Houwelingen. ISBN 978 90 377 0564 5
- Burgerperspectieven 2011 | 3* (2011). Paul Dekker en Josje den Ridder. ISBN 978 90 377 0582 9
- Advies over het Wmo-budget huishoudelijke hulp 2012* (2011). Ab van der Torre, Saskia Jansen en Evert Pommer. ISBN 978 90 377 0573 7 (webpublicatie)
- Oudere migranten . Kennis en kennislacunes* (2011). Maaike den Draak en Mirjam de Klerk. ISBN 978 90 377 0597 3 (webpublicatie)
- Armoedesignalement 2011* (2011). CBS/SCP. ISBN 978 90 357 1870 8
- Burgerperspectieven 2011 | 4* (2012). Josje den Ridder, Jeanet Kullberg en Paul Dekker. ISBN 978 90 377 0593 5
- Burgerperspectieven 2012 | 1* (2012). Paul Dekker, Josje den Ridder en Paul Schnabel. ISBN 978 90 377 0607 9
- Burgerperspectieven 2012 | 2* (2012). Josje den Ridder en Paul Dekker. ISBN 978 90 377 0617 8

