

Leeropbrengsten van scholen

Prof. Dr. R.J. Bosker
Drs. J. de Jong - Heeringa
GION, Instituut voor Onderzoek van Onderwijs
Rijksuniversiteit Groningen

Groningen, 26 september 2006

Inhoudsopgave

Managementsamenvatting	3
1. Inleiding	6
2. De huidige beoordelingspraktijk	8
3. De gewenste beoordelingspraktijk	10
4. Adviseren versus beoordelen	12
5. Een beknopte verkenning	13
5.1. Inleiding	13
5.2. Het 11-plus examen	13
5.3. No Child Left Behind	13
5.4. Toegevoegde waarde	14
5.5. Leerstandaarden	14
5.6. De Monitor Primair Onderwijs	15
6. Nadere analyse van enkele beoordelingsmethoden	17
7. Conclusies	19
Referenties	22

Managementsamenvatting

Recentelijk is er een discussie ontstaan over de toetsing van leerlingen aan het eind van het basisonderwijs. Deze discussie vloeit voort uit een drietal zaken.

In de eerste plaats constateert de Inspectie van het Onderwijs dat scholen die bij hun leerlingen de Eindtoets Basisonderwijs van de Cito-groep afnemen, nogal eens leerlingen uitsluiten van deelname. In de tweede plaats noemt de inspectie het probleem dat scholen weliswaar kunnen worden beoordeeld op hun opbrengsten, maar dat in de praktijk dit oordeel norm- en niet criterium-georiënteerd is. Met andere woorden: of de opbrengsten van een school als voldoende worden beoordeeld hangt af van de opbrengsten die andere scholen met een vergelijkbare leerlingpopulatie realiseren. In de derde plaats klinkt vanuit het afnemend vervolgonderwijs de roep om helderheid over het door de basisscholen bij hun leerlingen te realiseren eindniveau, zodat deze afnemende scholen meer duidelijkheid hebben over het niveau waarop ze kunnen voortbouwen.

Op grond van deze overwegingen heeft het Ministerie van OC&W de volgende hoofdvraag geformuleerd: *Moet de overheid andere instrumenten – dan de nu gebruikte schoolscores op de Cito-eindtoets – stimuleren of inzetten, gegeven de behoefte aan normering van leeropbrengsten?*

Centraal in dit advies staat het instrument en de methode om scholen te beoordelen, niet om individuele leerlingen te adviseren omtrent hun verdere loopbaan. De enige reden om aan dit laatste aandacht te besteden is dat thans een voor advisering bedoeld instrument (met name de Cito-eindtoets) voor de Inspectie van het Onderwijs het hulpmiddel bij uitstek is om de opbrengsten van scholen te beoordelen.

Om te kunnen bepalen of de opbrengsten die een school bij haar leerlingen realiseert aan de maat zijn, zijn heldere standaarden nodig. Deze standaarden moeten betrekking hebben op de vaardigheidsniveaus die leerlingen aan het einde van het basisonderwijs bereikt dienen te hebben conform de kerndoelen basisonderwijs. De kerndoelen geven domeinen aan, en het formuleren van standaarden vraagt om een operationele uitwerking en normering om te kunnen bepalen of een leerling in zo'n domein als voldoende vaardig kan worden aangemerkt. Om vervolgens te kunnen bepalen of een school voldoende opbrengsten bij haar leerlingen realiseert moet vervolgens een standaard ten aanzien van het percentage leerlingen worden geformuleerd dat aan de standaarden voldoet en/of er moet een standaard voor het gemiddelde beheersingsniveau worden geformuleerd. De standaard zou bijvoorbeeld kunnen luiden: 90 procent van de leerlingen behaalt tenminste het voldoende beheersingsniveau, en gemiddeld is het beheersingsniveau ook minimaal voldoende.

De toetsen die hiervoor worden ingezet moeten representatief zijn voor de kerndoelen. Ze hoeven niet volledig dekkend te zijn, zoals ook examenopgaven nooit volledig dekkend zijn voor het hele domein, maar daar wel een goede representatie van dienen te zijn.

Kan de Cito-eindtoets voor dit doel worden ingezet? Daar is *in principe* niets op tegen, met de kanttekening dat wanneer de school ook wordt afgerekend op de

scores die de leerlingen behalen er garanties dienen te komen over de naleving van afnameprotocollen. En bovendien moeten er heldere regels worden geformuleerd over het eventueel niet laten deelnemen van bepaalde leerlingen aan de toetsafname. Een leerachterstand van anderhalf jaar of meer hoeft daarbij geen bezwaar te zijn, omdat in dat geval een aangepaste toets (bijvoorbeeld de Cito-niveautoets) kan worden ingezet die het vaardigheidsniveau van de leerling op dezelfde schaal afbeeldt als de Cito-eindtoets. Het feit dat de Cito-eindtoets in eerste instantie vooral een adviesfunctie heeft, is niet strijdig met de functie van beoordeling van de opbrengsten van scholen. Sterker nog: het feit dat er voor de leerlingen iets op het spel staat wanneer ze de toets maken is alleen maar gunstig. Voorts is het juist wenselijk dat de voorspellende waarde van de Cito-eindtoets zeer goed is: dit impliceert dat hetgeen met deze toets wordt gemeten van belang is voor hetgeen er in het voortgezet onderwijs gebeurt.

Kunnen andere toetsen voor dit doel worden ingezet? Dat kan alleen als deze toetsen voldoen aan de psychometrische criteria van betrouwbaarheid en validiteit, zoals o.a. uitgewerkt door de Commissie Testaangelegenheden Nederland (COTAN) van het Nederlands Instituut van Psychologen (NIP). Voorts dienen de toetsen representatief te zijn voor de kerndoelen. Om deze redenen worden in de praktijk alleen de Cito-eindtoets en daarvan afgeleide varianten geaccepteerd door de Inspectie van het Onderwijs. Dat de inspectie in sommige gevallen ook de Cito-entreetoets betreft in de beoordeling van de opbrengstkwaliteit lijkt ons om twee redenen niet wenselijk. Ten eerste omdat deze toets mede tot doel heeft te diagnosticeren op welke gebieden er in jaargroep 8 nog onderwijskundig correctieve actie plaats zou kunnen of moeten vinden. Ten tweede omdat de afname van deze toets einde jaargroep 7 is, een jaar te vroeg dus.

Vanuit het oogpunt van belasting van scholen en leerlingen, en gelet op de rol die de Cito-eindtoets thans vervult (het werkt een 'high stakes' afnameconditie in de hand, dat wil zeggen er staat voor de leerlingen wat op het spel) is het ons inziens onwenselijk een additioneel instrument te ontwikkelen dat naast deze toets wordt ingezet, met als enig oogmerk de opbrengst van scholen in kaart te kunnen brengen.

Resteert de normeringsvraag. Het advies van de Onderwijsraad over de leerstandaarden, met name dat deel van het advies dat betrekking had op de afsluiting einde jaargroep 8, geeft een passend en adequaat antwoord op deze vraag. Weliswaar is dat advies opgesteld met het oog op het bevorderen van meer helderheid voor leerlingen en scholen met het oog op een soepel te verlopen onderwijsloopbaan van de individuele leerling, maar in de praktijk kan het eveneens voor de beoordeling van opbrengsten van scholen worden ingezet. Stel dat de Cito-eindtoets conform het voorgaande het instrument bij uitstek blijft om opbrengsten van scholen te beoordelen, in elk geval voor de domeinen taal, rekenen/wiskunde en wereldoriëntatie, zulks eventueel na een aanpassing om de toets meer recht te laten doen aan de kerndoelen. Dan kunnen vervolgens standaarden worden geformuleerd die de gewenste beheersingsniveaus weergeven. Zoals thans ranges in de Cito-eindtoetsscores worden aangegeven ter ondersteuning van de advisering en plaatsing van leerlingen, zo zouden ook zeer gedifferentieerd standaarden kunnen worden geformuleerd die passen bij de diverse trajecten in het voortgezet onderwijs.

Meer globaal zou men natuurlijk kunnen volstaan met de tweedeling zoals de Onderwijsraad die zelf voorstelde (minimum en voldoende) of conform een driedeling zoals we die toegepast wordt in de Verenigde Staten van Amerika (basis, voldoende en gevorderd). Vervolgens dient men dan op het niveau van scholen te specificeren aan welke standaard de gemiddelde opbrengsten moeten voldoen gegeven de samenstelling van de leerlingpopulatie en/of het aanvangsniveau van de leerlingen bij entree in de basisschool (of iets later).

De beslissing tot afname van de Cito-eindtoets (of een alternatieve toets) is aan de basisscholen zelf. Wil men basisscholen consequent en uniform op hun opbrengsten kunnen beoordelen, dan veronderstelt dit het volgende. Of alle basisscholen gebruiken de Cito-eindtoets op een standaard manier, en alle leerlingen, een enkele nadere te specificeren uitzondering daar gelaten, doen mee. Of de scholen die daaraan niet mee doen, nemen een toets af die betrouwbaar, valide en equivalent is. De bewijslast voor het voldoen aan deze eisen ligt bij de betreffende toetsontwikkelaars.

Een alternatief is een geheel andere route, namelijk die waarbij de overheid de scholen het gebruik van een dergelijke toets (in de praktijk zal dit dan de Cito-eindtoets zijn) dwingend voorschrijft. Dat kan de vorm krijgen van een verplichting die ertoe leidt dat de overheid de Cito-eindtoets inkoopt en dat alle scholen de toets bij al hun leerlingen volgens een standaard protocol afnemen. Of het kan de vorm krijgen van een nationaal in te stellen eindexamen basisonderwijs.

Het bezwaar dat men tegen deze gang van zaken zou kunnen uitvoeren zou de grondwettelijke vrijheid van onderwijs kunnen zijn. De reactie daarop is echter tamelijk direct: wie scholen de vrijheid gunt dient helder te zijn over de verwachte opbrengsten; wie het bezwaar valide vindt zou dan ook oppositie moeten voeren tegen de eindexamens in het voortgezet onderwijs.

Tenslotte: de hoofdvraag luidde of de overheid andere instrumenten – dan de nu gebruikte schoolscores op de Cito-eindtoets – moet stimuleren of inzetten, gegeven de behoefte aan normering van leeropbrengsten. Het antwoord daarop is ondubbelzinnig ontkennend: nee, dat is niet nodig.

1. Inleiding

Recentelijk is er een discussie ontstaan over de toetsing van leerlingen aan het eind van het basisonderwijs. Deze discussie vloeit voort uit een drietal zaken.

In de eerste plaats constateert de Inspectie van het Onderwijs dat scholen die bij hun leerlingen de Eindtoets Basisonderwijs van de Cito-groep afnemen, nogal eens leerlingen uitsluiten van deelname. Daar kunnen ze goede redenen voor hebben, zoals het voorkomen van onnodige teleurstellingen bij leerlingen met grote leerachterstanden. Maar het kan ook zijn dat scholen beducht zijn voor de “public-relations” effecten van een lage opbrengstscore, en een daaruit resulterende negatieve beoordeling door de inspectie van de gerealiseerde leeropbrengsten, en om die reden deelname door de zwakke presteerders beperken.

In de tweede plaats noemt de inspectie het probleem dat scholen weliswaar kunnen worden beoordeeld op hun opbrengsten, maar dat in de praktijk dit oordeel norm- en niet criterium-georiënteerd is. Met andere woorden: of de opbrengsten van een school als voldoende worden beoordeeld hangt af van de opbrengsten die andere scholen met een vergelijkbare leerlingpopulatie realiseren. Dit kan dus enerzijds leiden tot een situatie waarbij een school, absoluut gezien, leeropbrengsten op een adequaat niveau realiseert, maar relatief gezien, namelijk vergeleken met andere vergelijkbare scholen het slecht doet. Anderzijds kan het leiden tot een situatie waarbij de school onvoldoende leeropbrengsten realiseert maar het toch, namelijk vergeleken met andere vergelijkbare scholen, goed doet.

In de derde plaats klinkt vanuit het afnemend vervolgonderwijs de roep om helderheid over het door de basisscholen bij hun leerlingen te realiseren eindniveau, zodat deze afnemende scholen meer duidelijkheid hebben over het niveau waarop ze kunnen voortbouwen. De kerndoelen basisonderwijs komen hier enigszins aan tegemoet, maar er zijn geen operationele standaarden geformuleerd die duidelijk aangeven wanneer leerlingen wel of juist niet bepaalde kerndoelen geacht worden bereikt te hebben.

In de zogenaamde *governance*-besturingsfilosofie is een centraal element dat scholen een grote mate van autonomie wordt toegekend bij de inrichting en uitvoering van het primaire proces, onder de conditie van het voldoen aan deugdelijkheidseisen, maar dat tegelijkertijd scholen verantwoording dienen af te leggen over de gerealiseerde opbrengsten. Met andere woorden: de hierboven genoemde knelpunten zijn des te nijpender bij een dergelijke besturingsfilosofie.

Op grond van deze overwegingen heeft het Ministerie van OC&W de volgende hoofdvraag geformuleerd:

Moet de overheid andere instrumenten – dan de nu gebruikte schoolscores op de Cito-eindtoets – stimuleren of inzetten, gegeven de behoefte aan normering van leeropbrengsten?

Om deze hoofdvraag te kunnen beantwoorden, dienen in elk geval de volgende twee vragen aan bod te komen:

- a. Interfereert de advies-functie van de Cito-eindtoets met de beoordelingsfunctie? Zo ja, zijn er voor de beoordelingsfunctie dan alternatieven denkbaar?
- b. Hoe zou een normering van de – op een of andere wijze gemeten – opbrengsten van scholen eruit kunnen zien, zodat in absolute termen bepaald kan worden of deze opbrengsten aan de maat zijn?

Voor de duidelijkheid: Centraal in dit advies staat het instrument en de methode om scholen te beoordelen, niet om individuele leerlingen te adviseren omtrent hun verdere loopbaan. De enige reden om aan dit laatste aandacht te besteden is dat thans een voor advisering bedoeld instrument (met name de Cito-eindtoets) voor de Inspectie van het Onderwijs het hulpmiddel bij uitstek is om de opbrengsten van scholen te beoordelen.

2. De huidige beoordelingspraktijk

Aan het einde van het basisonderwijs nemen zo'n 85 procent van de basisscholen in Nederland de Cito-eindtoets bij hun leerlingen af. Ze doen dit met name in het kader van de ondersteuning bij de advisering en keuze van een geschikt vervoltraject in het voortgezet onderwijs. Voor de leerlingen die anderhalf of meer jaren leerachterstand hebben (zo'n vijf procent van de leerlingen), en voor wie de afname van de Cito-eindtoets weinig zinvol is, is de niveautoets basisonderwijs beschikbaar. Deze makkelijker toets rapporteert de resultaten op dezelfde schaal als de Cito-eindtoets. Maar bij slechts één procent (en niet de mogelijk vijf procent waarvoor deze toets bedoeld is) van de leerlingen wordt deze toets ook daadwerkelijk afgenomen. De helft van de overige vijftien procent scholen gebruikt andere toetsen die eveneens landelijk genormeerd zijn. Neemt een school geen Cito-eindtoets af, maar wel de Cito-entreetoets (die einde jaargroep 7 wordt afgenomen) of toetsen uit het Cito-leerlingvolgssysteem, dan gebruikt de inspectie deze gegevens. Dan is er nog een groep scholen die de eindtoets van het Gereformeerd Pedagogisch Centrum afneemt, een aangepaste versie van de Cito-eindtoets. Ca. 6 á 7 procent van de basisscholen neemt geen landelijk genormeerde toetsen aan het einde van het basisonderwijs af (Inspectie van het Onderwijs, 2003). De genoemde toetsgegevens gebruikt de inspectie om een oordeel uit te spreken over de leeropbrengsten van die scholen. Zij doet dit door de gemiddelde toetsgegevens van een school te vergelijken met die van, qua samenstelling van de leerlingpopulatie, vergelijkbare scholen.¹ Blijft de school drie jaar achtereen sterk² achter dan wordt het oordeel "onvoldoende" uitgesproken.

Het feit dat de inspectie deze toetsgegevens voor het uitspreken van een kwaliteitsoordeel gebruikt draagt enkele risico's in zich, en roept ook voorts vragen op. Het toetsen van leerlingen aan het einde van het basisonderwijs heeft als primaire functie het ondersteunen van de leerkracht bij het formuleren van een advies omtrent het vervoltraject danwel het verschaffen van informatie aan de school voor voortgezet onderwijs in het kader van het zogenoemde tweede toetsingsmiddel (het eerste is het advies). Op basis van het door de leerling gedemonstreerde vaardigheidsniveau kan een gefundeerde verwachting worden uitgesproken over de kans op succes in een bepaald vervoltraject. Aan een dergelijke toets stellen we als eis dat deze predictief valide is: op basis van de door leerling behaalde score kan, ceteris paribus, een redelijke voorspelling worden gedaan over het vervoltraject dat de leerling met succes zou kunnen voltooien. We merken het volgende op:

1. De school voor basisonderwijs hoeft een dergelijk toetsgegeven niet zelf te overleggen. De school voor voortgezet onderwijs moet haar plaatsingsbeslissing baseren op een dergelijk toetsingmiddel, althans als het gaat om plaatsing in vmbo-theoretisch, havo, of vwo. De bewijslast ligt dus bij de school voor voortgezet onderwijs.

¹ De scholen zijn daartoe in 7 groepen ingedeeld op basis van de percentages leerlingen met respectievelijk een 0,00-, 0,25- of 0,90-gewicht.

² "Sterk" wordt door de inspectie operationeel gedefinieerd als een halve standaarddeviatie of meer onder het gemiddelde van de vergelijkingsgroep.

2. De basisschool kan om voor haar moverende redenen leerlingen uitsluiten van toetsafname. Die reden kan gelegen zijn in het voorkomen van teleurstellende ervaringen bij leerlingen met name als voor de school zelf duidelijk is dat voor de vervolgkeuze (bijvoorbeeld praktijkonderwijs) een dergelijk toetsingsmiddel niet noodzakelijk is.
3. De basisschool kan echter ook op 'strategische' gronden leerlingen van deelname uitsluiten, namelijk als ze verwacht dat hiermee de gemiddelde resultaten van de school zodanig gedrukt worden dat dit in een negatieve beoordeling door de inspectie zou kunnen resulteren. Overigens zij opgemerkt dat, als daar dringende redenen voor zijn, de inspectie zelf ook toetsgegevens van bepaalde leerlingen buiten beschouwing laat bij het bepalen van de door de school gerealiseerde leeropbrengsten. De inspectie wil daarmee voorkomen dat scholen worden "gestraft" als ze met succes zorgleerlingen binnen boord weten te houden.³ De ca. 8,5 procent van de leerlingen die hun loopbaan in het leerwegondersteunend onderwijs zullen voortzetten worden door de inspectie echter wel meegenomen in de berekeningen.
4. Het uitsluiten van bepaalde leerlingen heeft voor hen tot nadelig gevolg dat ze het behaalde vaardigheidsniveau niet kunnen demonstreren, met als nadelige consequentie dat niet aan het licht kan komen dat een afgegeven advies er wel eens naast zou kunnen zitten.
5. Er zijn echter ook basisscholen die in jaargroep 8 een intelligentietoets bij hun leerlingen afnemen naast een Cito-toets (hetzij de eindtoets hetzij de entreetoets). Dit doen ze expliciet met het oogmerk eventueel onontdekt talent op te sporen, en ze wegen deze informatie mee wanneer het onderwijskundig rapport (in casu het advies) van de leerling wordt opgemaakt.

De huidige beoordelingspraktijk van de inspectie is dus gebaseerd op scores op toetsen die met een ander oogmerk bij leerlingen worden afgenomen. Vervolgens bekijkt de inspectie hoe de opbrengst die een school realiseert bij haar leerlingen zich verhoudt tot die van vergelijkbare scholen. Dit noemt men normgeoriënteerd beoordelen, dat wil zeggen dat men de resultaten afzet tegen die van een normpopulatie. Dit in tegenstelling tot criteriumgeoriënteerd beoordelen, waarbij de resultaten worden vergeleken met een vooraf geëxpliciteerd vaardigheidsniveau dat men als voldoende aanmerkt (van der Schoot, 2001). Zo'n criteriumgeoriënteerde benadering zou er bijvoorbeeld als volgt uit kunnen zien: de opbrengsten van de school worden als voldoende beoordeeld als de toetsscores van tenminste 95 procent van haar leerlingen aan een bepaalde standaard voldoen (bijvoorbeeld 530 of hoger op de Cito-eindtoets). Aan normgeoriënteerd beoordelen kleeft het mogelijke bezwaar dat een school relatief goed scoort maar absoluut niet. Of omgekeerd, dat de school absoluut gezien het goed doet, maar relatief niet.

³ De inspectie sluit leerlingen uit als ze te kort in Nederland zijn, of minder dan vier jaar op de betreffende basisschool onderwijs hebben genoten, of geïndiceerd zijn voor het praktijkonderwijs. Ook leerlingen die naar het vso of de expertisecentra gaan blijven buiten beschouwing. De laatste drie groepen omvatten in totaal ongeveer 2,1 procent van de leerlingen.

3. De gewenste beoordelingspraktijk

Sturing op output is een manier om een meer volgens marktprincipes ingericht systeem te coördineren. Het is ook een controlemechanisme bij uitstek in een land waar Artikel 23 van de Grondwet zo'n prominente rol speelt bij de inrichting van scholen en waar men voorts toe wil naar minder regels en meer autonomie voor onderwijsinstellingen. Sturen op output impliceert twee zaken: het formuleren van doelen, en het controleren of van doelbereiking sprake is. Het formuleren van doelen in het basisonderwijs krijgt gestalte in de kerndoelen. Controle op doelbereiking impliceert dus het nagaan of die kerndoelen, in meer of mindere mate, bereikt zijn.

De beoordeling van onderwijsopbrengsten vindt plaats op verschillende niveaus in het onderwijssysteem. De school beoordeelt de leerling, de inspectie de school, en voorts dient op nationaal niveau bewaakt te worden of de onderwijsresultaten aan de maat zijn. De school beoordeelt de leerling met behulp van proefwerken, schriftelijke en mondelinge overhoringen, methodegebonden en/of methodenonafhankelijke toetsen, toetsen uit leerlingvolgsystemen en eindtoetsen. De inspectie spreekt haar oordeel over de opbrengsten van de school uit op basis van de aggregatie van de scores die leerlingen behalen op de meer objectieve toetsen die de school zelf bij haar leerlingen afneemt. De kwaliteitsbewaking op systeemniveau vindt plaats door óf de diverse gegevens te aggregeren naar het nationale niveau, óf door kennis te nemen van de resultaten van specifiek daartoe ontworpen peilingen (gebaseerd op metingen bij leerlingen) van PPON of internationale assessment studies (van de IEA en OECD/PISA).

Het beoordelen van de door scholen bereikte leeropbrengsten veronderstelt het volgende:

1. de toetsen die de leerlingen maken vormen een goede afspiegeling van de kerndoelen;
2. de leerlingen zijn gemotiveerd voor het bereiken van een hoge toetsscore, met andere woorden "er moet iets op het spel staan";
3. de school is (tenminste gedeeltelijk) verantwoordelijk voor de door de leerlingen behaalde resultaten;
4. er is een standaard zodat duidelijk is wanneer een school voldoende leeropbrengsten heeft gerealiseerd.

ad 1) De Cito-eindtoets voldoet ten dele aan dit criterium. Met name de kerndoelen op het terrein van Nederlands, wiskunde en wereldoriëntatie worden redelijk afgedekt.⁴ Voor de overige domeinen (Friese taal, kunstzinnige oriëntatie, bewegingsonderwijs en, recentelijk, burgerschapsvorming) is dat echter niet het geval. Overigens zij opgemerkt dat een goede afspiegeling ook bereikt kan worden zonder dat volledigheid wordt nagestreefd en bereikt.

ad 2) Een toets die we bij een leerling zouden afnemen zonder dat de uitslag voor de leerling zelf gevolgen heeft, lokt uit dat de leerling niet naar vermogen zal presteren. Een toets met een adviserings- of een examenfunctie voldoet aan deze eis.

⁴ Anders dan bij de PPON-peilingen bevat de Cito-eindtoets echter geen zogenaamde authentieke taken.

ad 3) Dat de school verantwoordelijk is voor het bereikte resultaat is moeilijk voor honderd procent aannemelijk te maken. Leerlingen brengen maar 15 procent van hun “wakkere” tijd door op school, en leren ook in andere verbanden (zoals buurt, vrienden, gezin, familie). Om deze reden wordt vaak gepleit voor het bepalen van de door een school toegevoegde waarde in plaats van het waarderen van de door de school gerealiseerde leeropbrengsten in vergelijking met die van vergelijkbare scholen. De toegevoegde waarde is dan de gemiddelde vooruitgang die leerlingen boeken tijdens hun verblijf in de basisschool. In het volgende hoofdstuk gaan we nader op dit punt in.

ad 4) Op dit moment is de enige standaard die gehanteerd wordt een normgerelateerde: het gemiddelde niveau van vergelijkbare scholen minus een halve standaarddeviatie. Om toevalsfluctuaties te voorkomen wordt daarbij een termijn van drie jaren in beschouwing genomen.

Een gewenste beoordelingspraktijk is die waarbij alle scholen bij alle leerlingen onder een ‘high stakes’ situatie toetsen afnemen. De toetsen zijn afgeleid van de kerndoelen, en er is een normering vastgesteld voor zowel de leerlingen als de scholen (om te kunnen bepalen wat bijvoorbeeld een voldoende beheersingsniveau is). Om scholen recht te doen wordt op een of andere wijze bij de beoordeling betrokken met welk niveau de leerlingen de school zijn binnengekomen en/of de omstandigheden waarin de leerlingen buiten de school verkeren.

4. Adviseren versus beoordelen

We concludeerden aan het einde van het vorige hoofdstuk dat het beoordelen van scholen op basis van scores die leerlingen behalen op de Cito-eindtoets als voordeel heeft, dat leerlingen bij het maken van de toets hun best zullen doen omdat de advisering en dus hun toekomstige (school)loopbaan er mede van afhangt. Beoordelen van opbrengsten van scholen veronderstelt nu eenmaal dat de toetsen 'high stakes' zijn. Een soort examen aan het einde van het basisonderwijs zou in dat opzicht en binnen het kader van beoordeling van scholen ook een goede optie zijn.

De vraag is echter of de huidige praktijk waarin scholen door de Inspectie van het Onderwijs worden beoordeeld op basis van de Cito-eindtoets geen ongewenste effecten heeft c.q. zou kunnen hebben. Voorstelbaar is dat scholen mikken op goede schoolscores (een hoog gemiddelde op de Cito-eindtoets) in verband met die beoordeling. Dat kunnen ze deels bereiken door⁵

1. de minder presterende leerlingen uit te sluiten van deelname,
2. bij de afname van de Cito-eindtoets de leerlingen enigszins op weg te helpen, of
3. door bij het berekenen van de schoolscore de scores van bepaalde leerlingen buiten beschouwing te laten.

Zowel het eerste als tweede punt heeft gevolgen voor de adviseringsfunctie van de Cito-eindtoets. Door leerlingen uit te sluiten krijgen deze leerlingen geen kans om eventueel onontdekt talent te tonen. Dat heeft zowel voor die leerling gevolgen (uiteindelijk advisering en plaatsing zijn wellicht te laag), als ook zijn er op macro-niveau consequenties (rendementsverlies als een en ander later via omwegen gerepareerd moet worden, te laag gemiddeld scholingsniveau van de bevolking, etc.). Het op een of andere wijze (op weg) helpen van leerlingen bij de afname bedreigt de validiteit van de uitkomsten: de toets meet in dat geval niet meer sec wat leerlingen kunnen. De resulterende eindscore klopt dus niet, dat wil zeggen zal een overschatting zijn van het werkelijke vaardigheidsniveau van de leerling, en leidt dus tot adviseringsfouten.

Kortom: het gebruik van de Cito-eindtoets voor beoordelingsdoeleinden in een situatie waarbij scholen een zekere handelingsvrijheid hebben of nemen bij de afname, vormt een bedreiging voor de adviesfunctie. We laten het bij deze constatering in de wetenschap dat de Inspectie van het Onderwijs in haar rapport nader op deze kwesties ingaat, en vanwege het feit dat we ons hier concentreren op de centrale vraag: hoe kunnen de opbrengsten van scholen zo adequaat mogelijk worden beoordeeld? In dat licht merken we op dat het derde type strategisch gedrag van scholen (scores van leerlingen bij het berekenen van de schoolscore buiten beschouwing laten) eenvoudig te verhelpen is door derden die berekening uit te laten voeren.

⁵ In een aparte rapportage van de Inspectie van het Onderwijs wordt uitgebreider ingegaan op de vraag of en zo ja, hoe vaak deze zaken voorkomen.

5. Een beknopte verkenning

5.1. Inleiding

In dit hoofdstuk verkennen we enkele aanpalende onderwerpen die voor het onderhavige advies van belang zijn. Dat zijn twee interessante buitenlandse casussen die voor de Nederlandse situatie mogelijk leerzaam zijn. Het betreft enerzijds de Verenigde Staten van Amerika, waar sinds 2002 de No Child Left Behind wet duidelijke eisen stelt aan de door scholen te realiseren leeropbrengsten, en anderzijds het Verenigd Koninkrijk waar een 11-plus-examen bij leerlingen wordt afgenomen. Voorts staan we in dit hoofdstuk stil bij twee adviezen van de Onderwijsraad die aan het adviesthema raken: het advies over de toegevoegde waarde van een school en het advies over de leerstandaarden. Tenslotte gaan we kort in op het idee van een Monitor Primair Onderwijs, dat door sommigen, ten onrechte zoals zo zal blijken, wordt gezien als oplossing voor een aantal gesignaleerde problemen.

5.2. Het 11-plus examen

Het onderwijs in Engeland is verdeeld in 4 “key stages”, waarvan de tweede loopt van leerjaar 5 tot en met leerjaar 7, als leerlingen tussen de 7 en 11 jaar oud zijn. Elke key stage, dus ook deze tweede, wordt afgesloten met een nationale toets, die de vorm heeft van een mini-examen. De uitslag wordt uitgedrukt in een niveau van beheersing dat de leerling bereikt heeft, waarbij deze niveaus per “key stage” toenemen. Voor de afsluiting van de eerste periode wordt niveau twee als gemiddeld gezien. Zijn leerlingen 11 jaar oud, dan is het gebruikelijke niveau het vierde. De toetsuitslag heeft voor de leerling zelf geen directe consequenties (al was het maar omdat de leerlingen al weer ver gevorderd zijn in het volgende leerjaar voor de uitslag bekend wordt gemaakt), maar de door de leerlingen van een school gemiddeld behaalde uitslag wordt openbaar gemaakt. Een uitzondering op deze regel betreft enkele regio’s in Engeland, waar “grammar schools”, de scholen met een academische stroom die voorbereidt op de universiteit, de uitslag op het 11-plus examen gebruiken bij de selectie van leerlingen.

5.3. No Child Left Behind

In navolging van de staande succesvolle praktijk in de staten North-Carolina en Texas, is in de Verenigde Staten van Amerika in 2001 de No Child Left Behind wet van kracht geworden (Hess & Petrelli, 2006). Kernelementen zijn de volgende:

1. Elke staat formuleert zelf eenduidig voor de domeinen taal en rekenen drie beheersingsniveaus, namelijk “basis”, “voldoende” en “gevorderd”.
2. Om te bepalen over welke vaardigheid de leerlingen beschikken worden gestandaardiseerde toetsen afgenomen.

3. In 2013/2014 moet elke school bereikt hebben dat alle leerlingen op het niveau “voldoende” presteren, met uitzondering van maximaal 3 procent leerlingen die specifieke onderwijskundige zorg behoeven en/of de moedertaal maar beperkt machtig zijn.
4. Daartoe specificeert de school een traject waarlangs ze over de jaren bij het gestelde doel denkt aan te belanden voor de diverse onderscheiden groepen leerlingen (laag milieu, etnische minderheid, etc.).
5. Elk jaar wordt bepaald of de school voldoende “Annual Yearly Progress” boekt, namelijk door na te gaan of de school op koers ligt volgens het zelf uitgestippelde traject.
6. Er is een sanctie-ladder in geval de school jaarlijks onvoldoende vooruitgang boekt. Na het eerste jaar moet de school een verbeterplan opstellen en laten goedkeuren. Na het tweede jaar mogen leerlingen van school wisselen, en een school kiezen die wel voldoende vooruitgang boekt. Na het derde jaar dienen door de school gratis onderwijsfaciliteiten ingezet te worden om leerlingen die dat nodig hebben extra te ondersteunen. Na het vierde jaar volgt er correctieve actie (vervangen van een deel van het team, herziening van het curriculum, verlenging van de schooldag en/of het schooljaar). Na het vijfde jaar wordt de school grondig geherstructureerd.

5.4. Toegevoegde waarde

In 2003 bracht de Onderwijsraad het advies *Wat scholen toevoegen* uit. De kerngedachte van dit advies was, dat scholen publiekelijk verantwoording dienen af te leggen over de door hun gerealiseerde leeropbrengsten in termen van het meerdere dat de leerlingen op de school hebben geleerd. Anders gezegd: de gemiddelde leerwinst die leerlingen op een school boeken hoort aan de basis te staan van een beoordeling van de kwaliteit en niet de bruto opbrengsten aan het einde van jaargroep acht. Daarbij zouden tevens buitenschoolse invloeden (zoals gezinsomstandigheden) bij de bepaling van de toegevoegde waarde statistisch verrekend dienen te worden. Om logistieke redenen stelde de Onderwijsraad voor de leerwinst van de leerlingen te bepalen tussen jaargroep 3 en jaargroep 8. Een eerdere aanvangsmeting zou te veel complicaties geven. Andere studies van het Cito en de Inspectie van het Onderwijs lieten zien dat de door de Onderwijsraad voorgestelde benadering in de praktijk goed uitvoerbaar zou zijn, omdat veel scholen hun leerlingen reeds op jonge leeftijd toetsen – waarmee het bezwaar dat men jonge kinderen niet aan dergelijke ervaringen dient bloot te stellen door de onderwijspraktijk reeds achterhaald blijkt te zijn – , zodat het benodigde cijfermateriaal vaak reeds aanwezig is.

5.5. Leerstandaarden

Vier jaar eerder bracht de Onderwijsraad (1999a) het advies *Zeker weten, leerstandaarden als basis voor toegankelijkheid* uit. Dat advies is uitermate relevant met het oog op de normeringskwestie. De Onderwijsraad gaf in dat advies aan, dat er een operationele uitwerking van de kerndoelen, eindtermen en tussendoelen diende te komen, en in een tweetal aanvullende studies werd concreet toegelicht hoe dat zou kunnen (Onderwijsraad 1999b en c). Voorts werd aangegeven hoe men vervolgens duidelijke standaarden zou kunnen stellen op twee niveaus door gebruik te maken van procedures zoals die ook bij de PPON-peilingen worden toegepast:

1) *minimum*: datgene wat de leerling moet beheersen om wellicht met extra hulp en ondersteuning succesvol de loopbaan te kunnen vervolgen; deze standaard zou door bijna alle leerlingen bereikt moeten worden (de Onderwijsraad specificeerde dit als 90 procent of meer)

2) *voldoende*: leerlingen die aan deze standaard voldoen kunnen zonder meer succesvol hun loopbaan vervolgen (de Onderwijsraad specificeerde dit als 70 á 75 procent of meer).

De Onderwijsraad gaf voorts aan dat dergelijke leerstandaarden geformuleerd zouden moeten worden voor einde jaargroep 4, einde basisonderwijs en einde basisvorming. Voornaamste functie van de leerstandaarden is dat leerlingen die aan de standaarden voldoen in principe een onbelemmerde voortgang van hun onderwijs carrière wacht. Ze scheppen daarmee zowel helderheid voor het basisonderwijs (in termen van te behalen streefniveaus) als het voortgezet onderwijs (wat kunnen we als aanvangsniveau verwachten?).

5.6. De Monitor Primair Onderwijs

In de discussie met de Tweede Kamer over het onderhavige onderwerp is als mogelijke oplossing de Monitor Primair Onderwijs genoemd. De gedachten over dit onderwerp vloeiden voort uit twee wensen:

- a) de ontwikkeling van schoolspecifieke kwaliteitsmonitoring, en
- b) de ontwikkeling van een bovenschoolse, systeemgerichte kwaliteitsmonitoring.

Voor de tweede functie werd gedacht aan een meer permanente versie van de PPON-peilingen. Hierbij worden in de praktijk veel gebruikte toetsen gecombineerd met specifiek voor peilingsonderzoek ontwikkelde toetsen. Hierdoor wordt het eenvoudiger om de gegevens van een representatief deel van de scholen te gebruiken om op systeemniveau de onderwijskwaliteit (met name dan het onderwijsniveau) te kunnen bewaken. Interessant vanuit het perspectief dat individuele scholen beoordeeld zouden moeten worden op hun opbrengsten zijn de gedachten over de schoolspecifieke kwaliteitsmonitor. De gedachten komen erop neer dat op basis van de vele toetsen die scholen toch al gebruiken (met name de Cito-leerlingvolgsysteemtoetsen, de Entreetoets en de Eindtoets) zodanige rapportagevormen worden ontwikkeld dat scholen over opbrengstgerichte zelfevaluatiegegevens kunnen beschikken. Daarbij zouden dan ook criteriumgeoriënteerde (in plaats van normgeoriënteerde) beoordelingskaders en rapportages over toegevoegde waarde worden ontwikkeld. Deze functionaliteiten liggen dicht tegen het door ons in het voorgaande hoofdstuk wenselijk geachte beoordelingssysteem aan. Er zijn echter enkele wezenlijke verschillen. In de eerste plaats gaat het bij deze monitor om alle toetsen die scholen op enig moment in de basisschoolperiode afnemen, en dus niet alleen om opbrengstgegevens aan het

einde van de rit. In de tweede plaats is de klant of afnemer van dit product de school zelf, en niet de overheid of de namens die overheid opererende Inspectie van het Onderwijs. In operationele termen zou de uitwerking van de schoolspecifieke monitor op basis van de eindtoets basisonderwijs samenvallen met de wijze waarop de inspectie dergelijke opbrengstgegevens van scholen afleidt, alleen uitgebreid met een aantal extra rapportages zoals criteriumgeoriënteerde beoordelingskaders en rapportages over toegevoegde waarde tussen metingen. Het cruciale verschil met de PPON-peilingen is, dat de school in dit geval verantwoordelijk is voor de afname van toetsen en niet een overkoepelende instantie.

6. Nadere analyse van enkele beoordelingsmethoden

Wat zijn nu de overeenkomsten en verschillen tussen de verschillende methoden om de leeropbrengsten van scholen te beoordelen? En wat zijn voor- en nadelen van die methoden? En hoe uitvoerbaar zijn ze? Deze vragen staan in dit hoofdstuk centraal.

Grofweg kunnen we twee methoden onderscheiden:

1. het bepalen van de leeropbrengsten van scholen aan de hand van (leer)standaarden;
2. het bepalen van de toegevoegde waarde die scholen realiseren.

Beide methoden veronderstellen de afname van een toets bij leerlingen aan het einde van het basisonderwijs. De tweede methode vereist echter eveneens de afname van een toets op een eerder moment. En voor deze afname gelden dezelfde eisen als die we stellen aan de afname van een eindtoets: alle leerlingen moeten meedoen, de afnamecondities moeten gestandaardiseerd zijn en er moet iets op het spel staan voor de leerling. Het zal duidelijk zijn dat de laatste voorwaarde moeilijk te vervullen is, zij het dat in de praktijk jonge kinderen meestal wel intrinsiek gemotiveerd zijn voor het maken van een toets. Daarnaast is een aanvullende voorwaarde dat de begin- en eindmeting op dezelfde schaal geplaatst kunnen worden (we moeten immers de voortgang van leerlingen kunnen bepalen). De begin- en eindtoets moeten daartoe betrekking hebben op hetzelfde inhoudsdomen en de uitslagen moeten (door het inzetten van equivaleringmethoden) op elkaar herleidbaar zijn. Globaal genomen voldoen toetsen uit het Cito-leerlingvolgsysteem en de Cito-eindtoets aan deze voorwaarde.⁶

Wordt gekozen voor de methode van de toegevoegde waarde, dan is het ook nog mogelijk de methoden van de (leer)standaarden te hanteren: de vereiste gegevens zijn immers beschikbaar. Het omgekeerde geldt natuurlijk niet.

Toegevoegde waarde is gebaseerd op de voortgang die de leerlingen boeken. Het voordeel van deze methode is evident: scholen die te maken hebben met leerlingen met een bijzonder laag aanvangsniveau van de leerlingen worden daarvoor niet 'gestraft'. Het nadeel van deze methode springt echter ook meteen in het oog: de leerlingen kunnen vooruitgang boeken (en de school dus toegevoegde waarde realiseren) zonder dat de leerlingen een aanvaardbaar vaardigheidsniveau bereiken aan het einde van het basisonderwijs. Om deze reden kan de methode van het bepalen van de door scholen toegevoegde waarde nooit vervangend zijn voor het werken met een (leer)standaarden methode. Er is ook een praktisch motief de laatste methode te prefereren: het duurt zeker nog 7 jaren (we moeten leerlingen minimaal 6 jaren volgen en de implementatie van een en ander vereist ook nog minstens een jaar voorbereiding) voor de methode van de toegevoegde waarde bruikbare resultaten oplevert. Overigens is deze methode vanzelfsprekend arbeidsintensiever, omdat er een extra toets wordt afgenomen, een goede administratie moet worden bijgehouden, de begin- en eindmeting aan elkaar gekoppeld moeten worden, en dus kostbaarder. Ook komt de praktische uitvoerbaarheid in het gedrang in die gevallen

⁶ Het inzetten van bestaande toetsen voor het bepalen van de toegevoegde waarde kan echter weer strategisch gedrag van scholen uitlokken met nadelige gevolgen voor leerlingen analoog aan zoals we die eerder in hoofdstuk 4 beschreven voor de eindtoets.

waar veel leerlingen tussentijds van basisschool wisselen: aan welke school moet hun voortgang worden toegeschreven?

Is daarmee een criterium-georiënteerde benadering onder alle omstandigheden te prefereren? Het antwoord op die vraag is van twee omstandigheden afhankelijk: op hoeveel niveaus worden standaarden geformuleerd, en aan wie moet rekenschap worden afgelegd? Het is goed voorstelbaar dat veel scholen, als er slechts twee of drie niveau worden gehanteerd, hun leerlingen tot een voldoende of zelfs gevorderd niveau weten te brengen. Het werken met standaarden is dan niet erg onderscheidend, en diverse belangengroepen (ouders, schoolbesturen, maar ook de Inspectie van het Onderwijs en de overheid) hebben allicht behoefte aan meer gedifferentieerde informatie in het gesprek dat ze met de school wensen te voeren in het kader van *governance* (horizontale verantwoording) of toezicht van overheidswege (verticale verantwoording). In principe is de gewenste informatie wel beschikbaar in de vorm van toetsscores van leerlingen. Het werken met standaarden betekent alleen dat op basis van de “ruwe informatie” een waardering wordt gegeven.

Het werken met standaarden bij het beoordelen van de leeropbrengsten van scholen zal echter niet zonder slag of stoot kunnen worden geïntroduceerd. Alhoewel in het onderwijs evalueren aan de orde van de dag is (proefwerken, overhoringen, opstellen, toetsafnames) en waarderend dus ook (onvoldoendes en voldoende, de A-tot en met E-niveaus op de toetsen, overgaan of zittenblijven, advisering) is het niet gezegd dat scholen het zelf op prijs stellen geëvalueerd te worden op een criterium-georiënteerde wijze. De aanvaardbaarheid van een dergelijke beoordelingsmethode zal ons inziens toenemen als in het onderwijs de leerstandaarden geïntroduceerd zijn en scholen in hun dagelijkse praktijk zelf ervaren hoe een dergelijk criterium-georiënteerde beoordelingskader functioneert. Het voordeel van een dergelijke beoordelingsmethode boven het werken met toegevoegde waarde is bovendien dat een en ander beter navolgbaar en minder abstract is, hetgeen de acceptatie zal vergemakkelijken.

Tenslotte zij opgemerkt dat indien men de leeropbrengsten van scholen ook op domeinen als bijvoorbeeld kunstzinnige oriëntatie of burgerschapsvorming wil beoordelen, de methode van de standaarden aanmerkelijk minder gecompliceerd is dan de methode van de toegevoegde waarde. Zij het natuurlijk dat de Cito-eindtoets niet voorziet in de meting van deze domeinen, en er dus aanvullende instrumenten dienen te worden ontwikkeld.

7. Conclusies

Wat zijn nu de antwoorden op de gestelde vragen, en welk advies vloeit daaruit voort? Laten we eerst de twee deelvragen die we in deze studie trachten te beantwoorden recapituleren.

- a. Interfereert de advies-functie van de Cito-eindtoets met de beoordelingsfunctie? Zo ja, zijn er voor de beoordelingsfunctie dan alternatieven denkbaar?
- b. Hoe zou een normering van de – op een of andere wijze gemeten – opbrengsten van scholen eruit kunnen zien, zodat in absolute termen bepaald kan worden of deze opbrengsten aan de maat zijn?

Bij de beantwoording van deze vragen starten we met de laatste. Om te kunnen bepalen of de opbrengsten die een school bij haar leerlingen realiseert aan de maat zijn, zijn heldere standaarden nodig. Deze standaarden moeten betrekking hebben op de vaardigheidsniveaus die leerlingen aan het einde van het basisonderwijs bereikt dienen te hebben conform de kerndoelen basisonderwijs. De kerndoelen geven domeinen aan, en het formuleren van standaarden vraagt om een operationele uitwerking en normering om te kunnen bepalen of een leerling in zo'n domein als voldoende vaardig kan worden aangemerkt. Om vervolgens te kunnen bepalen of een school voldoende opbrengsten bij haar leerlingen realiseert moet vervolgens een standaard ten aanzien van het percentage leerlingen worden geformuleerd dat aan de standaarden voldoet en/of er moet een standaard voor het gemiddelde beheersingsniveau worden geformuleerd. De standaard zou bijvoorbeeld kunnen luiden: 90 procent van de leerlingen behaalt tenminste het voldoende beheersingsniveau, en gemiddeld is het beheersingsniveau ook minimaal voldoende.

De toetsen die hiervoor worden ingezet moeten representatief zijn voor de kerndoelen. Ze hoeven niet volledig dekkend te zijn, zoals ook examenopgaven nooit volledig dekkend zijn voor het hele domein, maar daar wel een goede representatie van dienen te zijn. Kan de Cito-eindtoets voor dit doel worden ingezet? Daar is *in principe* niets op tegen, met de kanttekening dat wanneer de school ook wordt afgerekend op de scores die de leerlingen behalen er garanties dienen te komen over de naleving van afnameprotocollen en er heldere regels worden geformuleerd over het eventueel niet laten deelnemen van bepaalde leerlingen aan de toetsafname. Zoals gesteld hoeft een leerachterstand van anderhalf jaar of meer geen bezwaar te zijn, omdat in dat geval een aangepaste toets (bijvoorbeeld de Cito-niveautoets) kan worden ingezet die het vaardigheidsniveau van de leerling op dezelfde schaal afbeeldt als de Cito-eindtoets. Het feit dat de Cito-eindtoets in eerste instantie vooral een adviesfunctie heeft, is niet strijdig met de functie van beoordeling van de opbrengsten van scholen. Sterker nog: het feit dat er voor de leerlingen iets op het spel staat wanneer ze de toets maken is alleen maar gunstig. Voorts is het juist wenselijk dat de predictieve validiteit van de Cito-eindtoets zeer goed is: dit impliceert dat hetgeen met deze toets wordt gemeten van belang is voor hetgeen er in het voortgezet onderwijs gebeurt. Daarmee hebben we de eerste vraag beantwoord.

Kunnen andere toetsen voor dit doel worden ingezet? Dat kan alleen als deze toetsen voldoen aan de psychometrische criteria van betrouwbaarheid en validiteit, zoals o.a. uitgewerkt door de Commissie Testaangelegenheden Nederland (COTAN) van het Nederlands Instituut van Psychologen (NIP). Voorts dienen de toetsen

representatief te zijn voor de kerndoelen. Om deze redenen worden in de praktijk alleen de Cito-eindtoets en de variant daarop van het Gereformeerd Pedagogisch Centrum geaccepteerd door de Inspectie van het Onderwijs. Dat de inspectie in sommige gevallen ook de Cito-entreetoets betreft in de beoordeling van de opbrengstkwaliteit lijkt ons om twee redenen niet wenselijk. Ten eerste omdat deze toets mede tot doel heeft te diagnosticeren op welke gebieden er in jaargroep 8 nog onderwijskundig correctieve actie plaats zou kunnen of moeten vinden. Ten tweede omdat de afname van deze toets einde jaargroep 7 is, een jaar te vroeg dus.⁷

Vanuit het oogpunt van belasting van scholen en leerlingen, en gelet op de rol die de Cito-eindtoets thans vervult (het werkt een 'high stakes' afnameconditie in de hand) is het ons inziens onwenselijk een additioneel instrument te ontwikkelen dat naast deze toets wordt ingezet, met als enig oogmerk de opbrengst van scholen in kaart te kunnen brengen. Weliswaar heeft het gebruik van de Cito-eindtoets voor de beoordeling van leeropbrengsten van scholen mogelijk (nadelige) gevolgen voor de adviesfunctie van die toets, maar het feit dat de Cito-eindtoets een adviesfunctie heeft heeft omgekeerd geen nadelige consequenties voor de beoordelingsfunctie.

Resteert de normeringsvraag. Het advies van de Onderwijsraad over de leerstandaarden, met name dat deel van het advies dat betrekking had op de afsluiting einde jaargroep 8, geeft een passend en adequaat antwoord op deze vraag. Weliswaar is dat advies opgesteld met het oog op het bevorderen van meer helderheid voor leerlingen en scholen met het oog op een soepel te verlopen onderwijsloopbaan van de individuele leerling, maar in de praktijk kan het eveneens voor de beoordeling van opbrengsten van scholen worden ingezet. Stel dat de Cito-eindtoets conform het voorgaande het instrument bij uitstek blijft om opbrengsten van scholen te beoordelen, in elk geval voor de domeinen Nederlandse taal, rekenen/wiskunde en wereldoriëntatie, zulks eventueel na een aanpassing om de toets meer recht te laten doen aan de kerndoelen. Dan kunnen vervolgens standaarden worden geformuleerd die de gewenste beheersingsniveaus weergeven.⁸ Zoals thans ranges in de Cito-eindtoetsscores worden aangegeven ter ondersteuning van de advisering en plaatsing van leerlingen, zo zouden ook zeer gedifferentieerd standaarden kunnen worden geformuleerd die passen bij de diverse trajecten in het voortgezet onderwijs. Meer globaal zou men natuurlijk kunnen volstaan met de tweedeling zoals de Onderwijsraad die zelf voorstelde (minimum en voldoende) of conform een driedeling zoals we die toegepast zagen worden in de Verenigde Staten van Amerika (basis, voldoende en gevorderd). Vervolgens dient men dan op het niveau van scholen te specificeren aan welke standaard de gemiddelde opbrengsten dienen te voldoen gegeven de samenstelling van de leerlingpopulatie en/of het aanvangsniveau van de leerlingen bij entree in de basisschool (of iets later).

⁷ Een speciaal geval vormen natuurlijk die scholen die de entreetoets in jaargroep 8 afnemen als alternatief voor de eindtoets. Maar in dit geval wordt de toets in de onderwijspraktijk niet conform de bedoelingen van de ontwikkelaar ingezet, en dat is derhalve ook niet wenselijk.

⁸ Voor een voorbeeldmatige uitwerking verwijzen we naar de bijlagen bij het advies van de Onderwijsraad *Zeker weten* (Onderwijsraad 1999b en c).

Resteert tenslotte nog een vraag. De beslissing tot afname van de Cito-eindtoets (of een alternatieve toets) is aan de basisscholen zelf. Wil men basisscholen consequent en uniform op hun opbrengsten kunnen beoordelen, dan veronderstelt dit het volgende. Of alle basisscholen gebruiken de Cito-eindtoets op een standaard manier, en alle leerlingen, een enkele nader te specificeren uitzondering daar gelaten, doen mee. Of de scholen die daaraan niet mee doen, nemen een toets af die betrouwbaar, valide en equivalent is. De bewijslast voor het voldoen aan deze eisen ligt bij de betreffende toetsontwikkelaars. Een alternatief is een geheel andere route, namelijk die waarbij de overheid de scholen het gebruik van een dergelijke toets (in de praktijk zal dit dan de Cito-eindtoets zijn) dwingend voorschrijft. Dat kan de vorm krijgen van een verplichting die ertoe leidt dat de overheid de Cito-eindtoets inkoopt en dat alle scholen de toets bij al hun leerlingen volgens een standaard protocol afnemen. Of het kan de vorm krijgen van een nationaal in te stellen eindexamen basisonderwijs.

Het bezwaar dat men tegen deze gang van zaken zou kunnen uitvoeren zou de grondwettelijke vrijheid van onderwijs kunnen zijn. De reactie daarop is echter tamelijk direct: wie scholen de vrijheid gunt dient helder te zijn over de verwachte opbrengsten; wie het bezwaar valide vindt zou dan ook oppositie moeten voeren tegen de eindexamens in het voortgezet onderwijs.

Bij het afwegen van de voor- en nadelen van het gebruik van de door scholen toegevoegde waarde respectievelijk het inzetten van een criterium-georiënteerd kader op basis van leerstandaarden voor het beoordelen van de opbrengsten van scholen, sloeg de balans door in het voordeel van de laatste methode.

Tenslotte: de hoofdvraag luidde of de overheid andere instrumenten – dan de nu gebruikte schoolscores op de Cito-eindtoets – moet stimuleren of inzetten, gegeven de behoefte aan normering van leeropbrengsten. Het antwoord daarop is ondubbelzinnig ontkennend: nee, dat is niet nodig.

Referenties

Blok, H., Otter, M., & Roeleveld, J. (2002). *Leerlingvolgsystemen: Praktische lessen uit onderzoek*. Amsterdam: SCO-Kohnstamm Instituut.

Hess, F.M., & Petrelli, M.J. (2006). *No Child Left Behind*. New York: Peter Lang.

Inspectie van het Onderwijs (2003). *Het bepalen van de toegevoegde waarde door basisscholen*. Utrecht: Inspectie van het Onderwijs.

Onderwijsraad (1999a). *Zeker weten. leerstandaarden als basis voor toegankelijkheid*. Den Haag: Onderwijsraad.

Onderwijsraad (1999b). *Leerstandaarden Nederlandse taal basisonderwijs*. Den Haag: Onderwijsraad.

Onderwijsraad (1999c). *Leerstandaarden rekenen en wiskunde basisonderwijs*. Den Haag: Onderwijsraad.

Onderwijsraad (2003). *Wat scholen toevoegen*. Den Haag: Onderwijsraad.

Schoot, F. van der (2001). *Standaarden voor kerndoelen basisonderwijs. De ontwikkeling van standaarden voor kerndoelen basisonderwijs op basis van resultaten uit peilingsonderzoek* (dissertatie UvA). Arnhem: Cito.

Staphorsius, G. (2006). *Naar een monitor primair onderwijs (interne notitie)*. Arnhem: Cito.