



Inspectie van het Onderwijs
*Ministerie van Onderwijs, Cultuur en
Wetenschap*

**Differentiatie van het
opleidingsniveau van ouders
bij het beoordelen van
opbrengsten van basisscholen**

18 oktober 2012

Inhoudsopgave

1 AANLEIDING TOT HET ONDERZOEK	9
1.1 FACTOREN DIE EEN ROL SPELEN BIJ OPBRENGSTEN VAN SCHOLEN.....	9
1.2 DE HUIDIGE CORRECTIEFACTOR: LEERLINGGEWICHT.....	10
1.3 ONDERZOEKSVRAAG.....	12
2 HUIDIGE SYSTEMATIEK BIJ DE BEOORDELING VAN OPBRENGSTEN VAN BASISSCHOLEN.....	15
2.1 CURVE MET JAARLIJKS GELIJKBLIJVENDE ONDERGRENZEN	15
2.2 ONDERSCHIED RISICOANALYSE EN BEOORDELING VAN SCHOLEN.....	17
3 ONDERZOEKSOPZET	19
3.1 BESCHIKBARE GEGEVENS VOOR DE SECUNDAIRE ANALYSES.....	19
3.2 ONDERZOCHE MODELLen VOOR HET CORRIGEREN VAN OPBRENGSTEN	20
3.3 ANALYSES.....	22
4 RESULTATEN VAN DE SECUNDAIRE ANALYSES	23
4.1 VERKLAARDE VARIANTIE	23
4.2 VERSCHILLEN IN BEOORDELING VAN SCHOLEN	24
4.3 EISEN AAN CORRECTIEFACTOREN EN MODELLen	29
5 CONCLUSIES EN DISCUSSIE	31

Samenvatting

Onderzoeksvraag en opzet van het onderzoek

De inspectie heeft op verzoek van de minister van Onderwijs, Cultuur en Wetenschap onderzocht of meer gedifferentieerde informatie over het opleidingsniveau van ouders tot een betere beoordeling kan leiden van de bijdrage van basisscholen aan de prestaties van leerlingen. Gegevens van de inspectie zelf en gegevens uit het COOL-cohortonderzoek zijn voor dit doel aan elkaar gekoppeld. Vervolgens zijn secundaire analyses uitgevoerd op een bestand van 347 COOL-scholen. De resultaten en conclusies zijn voorgelegd aan enkele wetenschappers, vertegenwoordigers van het ministerie van OCW en een werkgroep met vertegenwoordigers van de PO-Raad.

In deze rapportage onderzoekt de inspectie alleen of de grens tussen onvoldoende en voldoende opbrengsten op een andere wijze tot stand zou kunnen komen. De analyses zijn niet gedaan met het doel vast te stellen hoe excellente scholen opgespoord kunnen worden.

Het belang van correctiefactoren bij de beoordeling van opbrengsten

Jaarlijks gaat de inspectie door middel van een risicoanalyse voor alle basisscholen na of de gemiddelde opbrengsten aan het eind van de basisschoolperiode voldoende zijn. Als een school drie jaar achtereenvolgende onvoldoende presteert, doet de inspectie nader onderzoek en kan de school als zwak of zeer zwak beoordeeld worden. De norm voor voldoende presteren is niet voor alle scholen hetzelfde, maar varieert afhankelijk van de samenstelling van de leerlingbevolking. Naarmate scholen meer leerlingen hebben van wie de ouders relatief laag zijn opgeleid, wordt de norm waar ze aan moeten voldoen voor een voldoende beoordeling van de opbrengsten lager. De huidige inspectiesystematiek gebruikt het percentage leerlingen met een leerlinggewicht van 0,3 en 1,2 als correctiefactor bij de beoordeling van opbrengsten. De vergelijking van scholen wordt daardoor eerlijker, omdat op deze manier rekening gehouden wordt met de verschillen in leerlingbevolking. De inspectie gebruikt het leerlinggewicht om pragmatische redenen: er zijn landelijk simpelweg geen betere gegevens over het opleidingsniveau van ouders beschikbaar voor alle basisscholen.

Een beperking die inherent is aan het leerlinggewicht is, dat aan de hand hiervan alleen leerlingen opgespoord kunnen worden met relatief laagopgeleide ouders. Die leerlingen zijn niet evenwichtig over basisscholen verdeeld: er zijn scholen die veel gewogen leerlingen hebben, maar ook scholen die weinig of geen gewogen leerlingen in hun leerlingbevolking hebben. Zo heeft meer dan de helft van de basisscholen minder dan tien procent gewogen leerlingen. Binnen deze laatste groep basisscholen kan de inspectie geen nader onderscheid maken naar het opleidingsniveau van ouders. Scholen waar alle ouders hbo of wo hebben gevolgd, zijn daardoor niet te onderscheiden van scholen waar alle ouders een mbo-diploma hebben. De vergelijking van scholen zou dus nog eerlijker kunnen worden, als de inspectie over meer gedetailleerde gegevens zou beschikken over het opleidingsniveau van ouders. Het COOL-cohortonderzoek voorziet in zulke gegevens. De inspectie kon daardoor voor een steekproef van scholen beschikken over het percentage ouders met hooguit lager onderwijs, met een opleiding op lbo-niveau, met een opleiding met mbo-niveau en met een opleiding op het niveau van het hoger onderwijs.

Het is bekend dat leerlingen gemiddeld genomen beter presteren als hun ouders hoger zijn opgeleid, maar ook is bekend dat binnen een groep leerlingen met vergelijkbaar opgeleide ouders de autochtone leerlingen in het algemeen nog iets beter presteren dan allochtone leerlingen. De factor etniciteit zat vroeger verwerkt in de gewichtenregeling, maar in de nieuwe gewichtenregeling is dat niet langer het geval. De inspectie beschikt op basis van de onderwijsnummerbestanden wel over gegevens over de etniciteit van leerlingen en heeft deze gegevens daarom ook in de analyses betrokken.

Onderzochte modellen

De inspectie heeft zeven modellen van correctiefactoren onderzocht en de resultaten afgezet tegen de huidige beoordelingssystematiek, waarin alleen voor leerlinggewicht gecorrigeerd wordt. Het eerste model corrigeert voor leerlinggewicht en etniciteit, het tweede voor opleidingsniveau, het derde voor opleidingsniveau en etniciteit en het vierde voor zowel leerlinggewicht als opleidingsniveau en etniciteit. Het vijfde onderzochte model is een spiegeling van de huidige correctiefactor: in dit model is gecorrigeerd voor het percentage ouders met het opleidingsniveau hoger onderwijs op een school. Het zesde model gaat uit van zeven verschillende schoolgroepen, waarbij iedere schoolgroep getypeerd wordt door opleidingsniveau en etniciteit. In het zevende model zijn scholen met meer dan 10 procent gewogen leerlingen beoordeeld volgens de nu geldende systematiek, terwijl bij de overige scholen gecorrigeerd is voor het percentage ouders dat op het niveau van het hoger onderwijs is opgeleid.

Betere schatting van de bijdrage van scholen aan opbrengsten

Het percentage verklaarde variantie per model laat zien hoe sterk de correctie voor de factoren uit het model is. De huidige correctiefactor (het percentage gewogen leerlingen op schoolniveau) verklaart 48 procent van de opbrengstverschillen tussen scholen. De resterende verschillen kunnen toegeschreven worden aan andere kenmerken van leerlingen en aan de kwaliteit van het onderwijs op de scholen waar de leerlingen onderwijs volgen. Als naast het leerlinggewicht ook een meer gedifferentieerde indicator van het opleidingsniveau van ouders en etniciteit als correctiefactoren gebruikt worden, zoals in het vierde model gebeurd is, stijgt de verklaarde variantie naar 55 procent. Het toevoegen van meer en nauwkeuriger gegevens over leerlingkenmerken leidt dus tot een betere schatting van de bijdrage van scholen aan de prestaties die leerlingen behalen. De overige bovenbeschreven modellen halen die 55 procent niet.

Verschuiven van beoordelingen van opbrengsten van scholen

Vervolgens is nagegaan welke scholen binnen ieder model een voldoende of een onvoldoende beoordeling krijgen. Het percentage scholen dat drie jaar achtereen onvoldoende presteert is bij alle onderzochte modellen steeds redelijk vergelijkbaar; dit varieert tussen 6,3 en 8,2 procent van de steekproef. Voor 96 procent van alle steekproefscholen maakt het gebruikte model van correctiefactoren niets uit: deze scholen worden consequent in alle modellen als voldoende of onvoldoende aangemerkt. Voor de resterende scholen maakt het gebruikte model van correctiefactoren wel iets uit. In model 4, dat de meeste variantie verklaart, krijgt 3 procent van de scholen een andere beoordeling dan in de huidige inspectiesystematiek. In de steekproef van 347 scholen verschuiven 6 scholen van een onvoldoende naar een voldoende beoordeling, terwijl 4 scholen van een voldoende naar een onvoldoende verschuiven. Landelijk zou dit neer kunnen komen op verschuivingen van beoordelingen bij zo'n 210 van de circa 7.000 basisscholen. Deze schatting moet met de nodige voorzichtigheid bekeken worden, omdat de onderzochte COOL-steekproef niet representatief is: scholen met veel

achterstandsleerlingen zijn oververtegenwoordigd en scholen met minder dan tien leerlingen in groep 8 zijn niet in de analyses betrokken.

Het verschuiven van de beoordeling van opbrengsten van scholen in de analyses betekent nog niet automatisch dat een deel van de huidige zwakke of zeer zwakke scholen bij andere correctiefactoren een andere beoordeling zou krijgen, of dat scholen die nu basistoezicht hebben automatisch zwak of zeer zwak zouden worden. Het oordeel zwak of zeer zwak wordt namelijk nooit alleen op de opbrengsten gebaseerd, maar vergt een diepergaand onderzoek bij besturen en scholen. Pas op basis van zo'n onderzoek kunnen scholen als zwak of zeer zwak beoordeeld worden.

Typering van verschuivende scholen

Critici van de huidige beoordelingssystematiek van de inspectie vermoeden doorgaans dat scholen met veel gewogen leerlingen eerder voldoende opbrengsten zouden halen bij een systematiek die voor meer factoren corrigeert, terwijl scholen met veel leerlingen die hoger opgeleide ouders hebben eerder een onvoldoende beoordeling zouden krijgen. Deze verwachting wordt door de analyses niet consistent ondersteund. De scholen die in het meest verklarende model van onvoldoende naar voldoende verschuiven of omgekeerd vormen een heterogene groep. Het is dus niet zo dat alleen scholen met veel gewogen leerlingen naar een voldoende verschuiven en het is evenmin zo dat alleen scholen met veel hoogopgeleide ouders naar een onvoldoende verschuiven. De verschuivende scholen verschillen onderling qua leerlingbevolking. Een overeenkomst voor een deel van deze scholen is wel dat ze relatief veel niet-westerse allochtone leerlingen hebben (de scholen die naar een voldoende gaan) of juist relatief weinig (de scholen die naar een onvoldoende gaan).

Consequenties van eventuele veranderingen

Voor een betere schatting van de bijdrage die scholen leveren aan de prestaties van hun leerlingen zijn dus in ieder geval meer gegevens nodig over etniciteit en opleidingsniveau. De etniciteit van leerlingen zat vroeger in het leerlinggewicht verwerkt, maar dat is niet meer zo. Toch zouden scholen niet extra bevraagd hoeven te worden om gegevens over etniciteit te verkrijgen, omdat deze al beschikbaar zijn via het onderwijsnummer.

Voor het opleidingsniveau van ouders ligt dat heel anders. Een substantiële verbetering van correctiefactoren is alleen mogelijk als de inspectie meer gedetailleerde gegevens over het opleidingsniveau van beide ouders per leerling krijgt binnen het kader van een formele regeling waar alle scholen onder vallen, vergelijkbaar met de huidige gewichtenregeling. Een formele regeling is nodig om sancties te kunnen treffen als scholen geen of foute gegevens verstrekken. De mate van detaillering van opleidingsgegevens zou om inhoudelijke redenen nog groter moeten zijn dan wat het COOL-cohort op dit moment kan bieden. Zo is een nader onderscheid van opleidingsniveaus binnen het vmbo wenselijk, evenals een nader onderscheid in het mbo en het hoger onderwijs. Bovendien zou vervolgens een nieuwe maat ontwikkeld moeten worden die op schoolniveau weergeeft hoe de leerlingbevolking in elkaar zit. Deze gegevensverzameling leidt tot een aanzienlijk zwaardere administratieve belasting van scholen dan de huidige gewichtenregeling. De inspectie acht een dergelijke exercitie bovendien alleen zinvol binnen het kader van een verplichte identieke eindtoets voor alle basisscholen, omdat alleen dan alle scholen met dezelfde maat gemeten kunnen worden. Dat kader is nog niet gewaarborgd.

Nadelen van veranderingen in de huidige inspectiesystematiek

Veranderen van de huidige inspectiesystematiek heeft ook nadelen. De huidige systematiek is transparant voor scholen, omdat ze aan hun percentage gewogen

leerlingen genoeg hebben om vast te stellen welke gemiddelde score ze minimaal op de Eindtoets Basisonderwijs moeten halen voor een voldoende beoordeling. Bovendien kunnen ze gemakkelijk een indruk krijgen van de prestaties van scholen die qua leerlingbevolking sterk op hen lijken. Deze transparantie bevordert het opbrengstgericht werken door scholen. Naarmate het aantal correctiefactoren voor de opbrengsten toeneemt, neemt de transparantie van de beoordelingssystematiek af en daarmee de betekenis van de opbrengstennormen als sturingsinstrument voor scholen die opbrengstgericht willen werken.

Een tweede nadeel van veranderingen op dit moment is, dat op voorhand bekend is dat de geschetste systematiek wel iets beter corrigeert dan de huidige, maar nog zeker niet optimaal. Van betere schattingen van de zuivere bijdrage van scholen aan opbrengsten kan pas sprake zijn als met modellen voor leerwinst gewerkt kan worden. Dat kan nu nog niet, omdat er voor het begin van de basisschool of voor momenten tijdens de schoolloopbaan nog geen toetsen zijn die inhoudelijk goed vergelijkbaar zijn met de Eindtoets. Voor goede analyses inzake de bijdrage van scholen aan opbrengsten is het verder bovendien zeer gewenst dat met andere analysetechnieken (zoals meerniveau-analyse) gewerkt kan worden, maar de inspectie moet dan wel kunnen beschikken over gegevens op het niveau van individuele leerlingen. Nu zijn voor de inspectie alleen gegevens op schoolniveau beschikbaar.

Ook bij leerwinstmodellen zijn overigens meer gedetailleerde opleidingsgegevens van ouders wenselijk en op termijn is een aanpassing van de gegevensbevraging bij scholen dan ook gewenst.

Conclusie

Veranderen van beoordelingssystematiek heeft het voordeel dat een enigszins betere schatting van de bijdrage van basisscholen aan prestaties mogelijk wordt. De inspectie vindt dit voordeel echter onvoldoende opwegen tegen de zwaardere bevragingslast voor scholen en het verlies aan transparantie van de beoordelingssystematiek. De nadelen wegen des te zwaarder omdat de verfijningen die nu haalbaar zijn, niet meteen leiden tot een optimaal model. Daarvoor zijn meer ontwikkelingen nodig op het terrein van leerwinstbepaling. De komende jaren valt daar waarschijnlijk het nodige van te verwachten.

De inspectie geeft er daarom de voorkeur aan de huidige werkwijze te handhaven, evenals de huidige beslisregels rond de beoordeling van de opbrengsten van basisscholen. Als besturen echter vinden dat de beslisregels voor een specifieke school niet adequaat zijn, bijvoorbeeld omdat van bijzondere of ongewone omstandigheden in de leerlingbevolking sprake is, kunnen zij dit in een gesprek met de inspectie beargumenteren en nader verantwoorden. Inspecteurs hebben altijd de professionele vrijheid om, waar dat nodig en passend is, beredeneerd af te wijken van de beslisregels. De inspectie prefereert deze professionele uitwisseling van argumenten tussen besturen en inspecteurs boven het nu aanpassen van de beoordelingssystematiek, gezien de bovenbeschreven nadelen.

De betrokken externe deskundigen en de PO-raad hebben aangegeven de conclusie te steunen dat een nadere differentiatie van het opleidingsniveau van ouders weliswaar een verfijning betekent van de beoordeling van opbrengsten, maar dat deze verbetering niet opweegt tegen de extra inspanning die hiervoor van scholen moet worden gevraagd.

1 Aanleiding tot het onderzoek

Voor basisscholen is de wijze waarop hun opbrengsten worden beoordeeld door de inspectie van groot belang, omdat een onvoldoende beoordeling er toe kan leiden dat een school als zwak of zeer zwak getypeerd wordt (dat dit zeker geen automatisme is, wordt in het volgende hoofdstuk van dit rapport toegelicht). Voor de acceptatie van de opbrengstenbeoordeling door basisscholen is het belangrijk dat de inspectiesystematiek voldoende rekening houdt met verschillen tussen scholen die te maken hebben met de achtergrond van leerlingen, zoals het opleidingsniveau van ouders. Een leerlingkenmerk als het opleidingsniveau van de ouders is door scholen niet te beïnvloeden, terwijl het wel van invloed is op de prestaties. Als de inspectiesystematiek voldoende rekening houdt met de invloed van deze door scholen onbeïnvloedbare leerlingkenmerken, worden de opbrengsten van scholen op een eerlijke manier met elkaar vergeleken.

Met enige regelmaat worden vragen gesteld over de beoordelingssystematiek van de inspectie. Die vragen zijn afkomstig uit het onderwijsveld (o.a. Godlieb, 2008), maar ook uit de politiek. Zo stelden de kamerleden Dijsselbloem en Smits tijdens het bespreken van de gewijzigde Wet op het Onderwijstoezicht in 2011 vragen over de correcties voor het opleidingsniveau van ouders die de inspectie toepast. De minister van Onderwijs, Cultuur en Wetenschap heeft de inspectie vervolgens verzocht door middel van secundaire analyses na te gaan welke mogelijkheden tot verbetering van de huidige inspectiesystematiek er zijn en tot wat voor administratieve belasting van scholen een eventuele andere werkwijze zou leiden (OCW, 2011a, 2012a). Dit rapport bevat de resultaten van de secundaire analyses die de inspectie heeft uitgevoerd.

In dit hoofdstuk komt eerst kort aan de orde welke factoren een rol spelen bij het tot stand komen van opbrengsten van scholen en wat dit betekent voor het schatten van de bijdrage van schoolse factoren (1.1). Vervolgens wordt de huidige correctiefactor die de inspectie gebruikt, het leerlinggewicht, beschreven (1.2). Ten slotte wordt de onderzoeksvraag voor de secundaire analyses geformuleerd (1.3).

1.1 Factoren die een rol spelen bij opbrengsten van scholen

De ontwikkeling van leerlingprestaties komt tot stand door een samenspel van verschillende factoren:

- De kwaliteit van het onderwijs op de basisschool,
- Andere schoolse factoren (zoals bijvoorbeeld de leerlingpopulatie van de basisschool),
- Aanleg/intelligentie en motivatie van leerlingen,
- Invloeden vanuit het gezin (zoals bijvoorbeeld de taal die thuis gesproken wordt),
- Overige buitenschoolse factoren.

Voor de inspectie gaat het bij de beoordeling van opbrengsten om het vaststellen van de bijdrage die scholen leveren aan prestaties, ofwel de toegevoegde waarde van scholen. Bij het meten van deze toegevoegde waarde van een school gaat het er om zo goed mogelijk in beeld te krijgen wat de bijdrage van de kwaliteit van het geboden onderwijs is, naast de bijdragen van de andere genoemde factoren. Om dat te bepalen is het nodig te corrigeren voor die andere factoren.

Leerwinst

Idealiter wordt de toegevoegde waarde van de school vastgesteld op basis van de leerwinst van individuele leerlingen, ofwel de prestatiegroei tussen twee of meer toetsmomenten, gemeten met toetsen die met elkaar vergeleken kunnen worden. Is de leerwinst hoger dan te verwachten is, dan wijst dat in de richting van een relatief hoge toegevoegde waarde van de school. Valt de leerwinst lager uit dan te verwachten is, dan is de bijdrage van de school aan de lage kant. Uiteraard moeten vervolgens ook leerlingkenmerken (zoals aanleg/intelligentie en motivatie) en buitenschoolse invloeden zoveel mogelijk verdisconteerd worden, omdat deze factoren ook een rol spelen bij de behaalde leerwinst. De kans dat de zuivere bijdrage van de school goed wordt geschat is op basis van leerwinst echter redelijk hoog.

Leerwinstbepalingen en het bepalen van de bijdrage van scholen aan prestaties op basis daarvan zijn nu nog niet mogelijk, omdat er aan het begin van de basisschool of op tussenmomenten in de schoolloopbaan nog geen toetsen zijn die zich inhoudelijk goed laten vergelijken met de Eindtoets in groep 8. Er lopen wel pilots op het gebied van leerwinstbepaling: hiervan worden in 2013 de eerste resultaten verwacht (OCW, 2011b). Op de langere termijn wordt het wellicht mogelijk om de beoordeling van opbrengsten van basisscholen op leerwinst te baseren.

Prestaties aan het eind van de basisschool

Gezien deze situatie is het voor dit moment nodig om te zoeken naar de 'next best' oplossing. Die ligt in het corrigeren van de prestaties aan het eind van de basisschool voor zoveel mogelijk in principe niet door het onderwijs beïnvloedbare factoren, zoals de aanleg/intelligentie en buitenschoolse factoren. Wat de buitenschoolse factoren betreft, is uit onderzoek gebleken dat veel verschillende factoren een rol kunnen spelen. Als over al die factoren gegevens verzameld moeten worden bij alle scholen, levert dat een hoge administratieve belasting op die niet aantrekkelijk is. Daarom is beperking tot de best voorspellende factoren wenselijk. Dat zijn het opleidingsniveau en de etnische herkomst van ouders van leerlingen; het inkomen van ouders hoort niet bij de beste voorspellers (Roeleveld, Mooij, Fettelaar en Ledoux 2011).

1.2 De huidige correctiefactor: leerlinggewicht

Toekennen van gewichten

Bij de huidige correctie maakt de inspectie om pragmatische redenen gebruik van het leerlinggewicht dat aan leerlingen van basisscholen wordt toegekend, als benadering van het opleidingsniveau van de ouders; er zijn geen andere gegevens voor alle scholen beschikbaar. De gewichten zijn voor alle leerlingen en daardoor voor alle scholen bekend en dat maakt het vergelijken van scholen op dit kenmerk mogelijk. Bij de gewichtenregeling worden ouders eerst in drie onderscheiden categorieën ingedeeld:

- Categorie 1: maximaal (speciaal) basisonderwijs of (v)so-zmlk
- Categorie 2: maximaal lbo/vbo, praktijkonderwijs of vmbo-basis- of kaderberoepsgerichte leerweg
- Categorie 3: overig voortgezet onderwijs en hoger.

Op basis hiervan krijgen leerlingen vervolgens een gewicht toegekend:

- Het gewicht 1,2 krijgen leerlingen van wie één van de ouders een opleiding heeft gehad uit categorie 1 en de andere ouder een opleiding uit categorie 1 of 2.
- Het gewicht 0,3 krijgen leerlingen van wie de ouders een opleiding uit categorie 2 hebben gehad.

- Het gewicht 0 krijgen alle leerlingen van wie één van de ouders een opleiding heeft gehad uit categorie 3 (CFI, 2008).

Van alle leerlingen in het basisonderwijs heeft 13 procent een gewicht 0,3 of 1,2 (OCW, 2011, Kerncijfers). Vroeger werd ook de etniciteit van leerlingen verwerkt in het bepalen van gewichten; bij de herziening van de gewichtenregeling in 2006 is die factor echter verdwenen.

Ongelijke verdeling gewogen leerlingen over scholen

Voor scholen die veel gewogen leerlingen hebben, ligt de ondergrens voor wat volgens de inspectie nog voldoende presteren is, lager dan voor scholen die weinig of geen gewogen leerlingen hebben. Voor een school met 90 procent gewogen leerlingen ligt de ondergrens voor een voldoende beoordeling van de score op de Eindtoets Basisonderwijs bijvoorbeeld rond 527, voor een school met 10 procent gewogen leerlingen rond 534. Zou deze correctie niet worden toegepast, dan zou de norm voor voldoende presteren voor alle scholen hetzelfde zijn. Dat zou alleen reëel zijn, als alle scholen ongeveer evenveel gewogen leerlingen zouden hebben. Dat is echter niet het geval: sommige scholen hebben zeer veel gewogen leerlingen, andere haast geen enkele (tabel 1).

Tabel 1 Verdeling van scholen naar percentage gewogen leerlingen, peildatum 1 oktober 2011 (absolute aantallen en percentages scholen)

Percentage gewogen leerlingen	Aantal scholen	Percentage scholen
0	544	8
1-5	1.907	28
6-10	1.559	23
11-15	923	14
16-20	575	8
21-30	564	8
31-40	283	4
41-50	182	3
51-60	156	2
61-70	79	1
71-80	30	<1
81-90	6	<1
91-100	0	0
Totaal	6.808	100

Bron: Inspectie van het Onderwijs, 2012

Slechts 14 procent van de basisscholen komt qua leerlingsamenstelling redelijk overeen met de landelijke percentages van gewogen en ongewogen leerlingen: deze groep scholen heeft tussen 11 en 15 procent gewogen leerlingen. Een kwart van de scholen (27 procent) heeft meer gewogen leerlingen dan op grond van de landelijke cijfers verwacht kan worden, terwijl meer dan de helft van de scholen (59 procent) daarentegen minder gewogen leerlingen heeft of zelfs geen enkele.

Beperkingen van het leerlinggewicht als correctiefactor

De gewichtenregeling definieert vooral de ouders met een laag opleidingsniveau. Dit is een serieuze beperking voor een goede correctie, omdat bekend is dat kinderen van universitair opgeleide ouders gemiddeld genomen beter presteren dan kinderen van hbo-opgeleide ouders, die het weer beter doen dan kinderen van mbo-opgeleide ouders (o.a. Roeleveld, 2003; Onderwijsraad, 2003). Binnen al deze

opleidingscategorieën presteren allochtone leerlingen lager dan autochtone leerlingen, ook als hun ouders een vergelijkbaar opleidingsniveau hebben (Driessen, Mulder en Roeleveld, 2012). De etnische factor speelt dus naast de opleiding van ouders ook nog een rol. Voor de ingang van de nieuwe gewichtenregeling maakte de factor etniciteit deel uit van het leerlinggewicht, nu is dat niet meer het geval.

Betere inschatting wenselijk

In de grote groep scholen die weinig gewogen leerlingen hebben, is het met de beschikbare gegevens voor de inspectie niet mogelijk om een nader onderscheid naar opleidingsniveau van ouders te maken. Scholen waar alle ouders hbo of wo hebben gevolgd, zijn daardoor niet te onderscheiden van scholen waar alle ouders een mbo-diploma hebben. De correctie op het opleidingsniveau van ouders door het leerlinggewicht kan hierdoor onvoldoende recht doen aan reële verschillen in de leerlingbevolking. Dat kan scholen benadelen die veel leerlingen hebben van wie de ouders qua opleiding dicht tegen de criteria voor de gewichtenregeling aanleunen, terwijl scholen met leerlingen van wie de ouders (zeer) hoog zijn opgeleid wellicht te gemakkelijk als voldoende presterend worden gezien. Bij de eerste groep scholen wordt hierdoor wellicht te vaak gesteld dat de prestaties onvoldoende zijn, bij de tweede groep gebeurt dat dan juist te weinig. Analyses wijzen uit dat meer informatie over de opleiding van ouders voor een deel van de scholen kan leiden tot andere beoordelingen (Roeleveld, Mooij, Fettelaar en Ledoux, 2011).

1.3 Onderzoeksvraag

Op verzoek van OCW heeft de inspectie door middel van secundaire analyses de volgende vraag onderzocht:

Welke effecten heeft het gebruik van meer gedifferentieerde informatie over het opleidingsniveau van ouders op de beoordeling van prestaties van scholen op de Eindtoets Basisonderwijs, in vergelijking met de huidige inspectiesystematiek?

Bij dit onderzoek blijft de principiële vraag of het corrigeren voor opleidingsniveau van ouders überhaupt gewenst is buiten beschouwing. Tegenstanders van deze correctie wijzen er op dat het corrigeren van opbrengsten er toe leidt dat de maatschappij van sommige scholen lagere prestaties accepteert dan van andere. In dit verband wordt er ook op gewezen dat scholen met veel gewogen leerlingen meer financiering krijgen dan scholen zonder deze leerlingen, waarbij de vraag gesteld wordt tot welke prestaties deze financiering eigenlijk minimaal zou moeten leiden. Dit type vragen valt buiten het bestek van deze rapportage. De uitgevoerde analyses hebben daar ook geen betrekking op.

Bij de beoordeling van opbrengsten van scholen is de afgelopen jaren veel aandacht uitgegaan naar het opsporen van zwakke en zeer zwakke scholen. Inmiddels is maatschappelijk ook meer aandacht ontstaan voor scholen die mogelijk excellent presteren en werkt OCW aan procedures om dergelijke scholen op te sporen. De hier gerapporteerde analyses hebben niet met die procedures te maken. Vergelijkbare analyses kunnen wellicht bruikbaar zijn om excellente scholen te traceren, maar in deze rapportage gaat het steeds alleen om het bepalen van de grens tussen onvoldoende en voldoende opbrengsten.

De prestaties van leerlingen zijn niet de enig denkbare indicator voor het beoordelen van de opbrengsten van scholen. Ook het percentage zittenblijvers, het percentage verwijzingen naar het speciaal onderwijs, het percentage leerlingen dat vooruitloopt op leeftijdgenoten, de spreiding tussen prestaties van leerlingen en het compenserend vermogen van de school voor de herkomst van leerlingen kunnen als

indicatoren beschouwd worden. Op dit moment vormen de prestaties voor taal en rekenen/wiskunde, zoals gemeten met toetsen in groep 8, wel de belangrijkste indicator van de opbrengsten van scholen in de risico-analyse die de inspectie jaarlijks voor alle basisscholen uitvoert. In deze rapportage gaat het daarom ook alleen om de vraag of die indicator beter geschat zou kunnen worden dan nu gebeurt.

De resultaten en conclusies zijn voorgelegd aan vertegenwoordigers van OCW, aan een werkgroep bestaande uit leden van de PO Raad en aan enkele wetenschappers (zie bijlage 1).

2 Huidige systematiek bij de beoordeling van opbrengsten van basisscholen

De wijze waarop de inspectie de opbrengsten aan het eind van het basisonderwijs beoordeelt, is met ingang van het schooljaar 2011/2012 veranderd. De nieuwe systematiek wordt in dit hoofdstuk beschreven (2.1), evenals het doel van de opbrengstenanalyse. De inspectie gebruikt de beoordeling van opbrengsten om risico's op scholen te signaleren: als een school onvoldoende opbrengsten heeft, bespreekt de inspectie deze bevindingen met het bestuur en kan nader onderzoek op de school in kwestie plaatsvinden. Pas na een dergelijk onderzoek komt een beoordeling als zwakke of zeer zwakke school in beeld. Deze werkwijze wordt eveneens kort toegelicht (2.2).

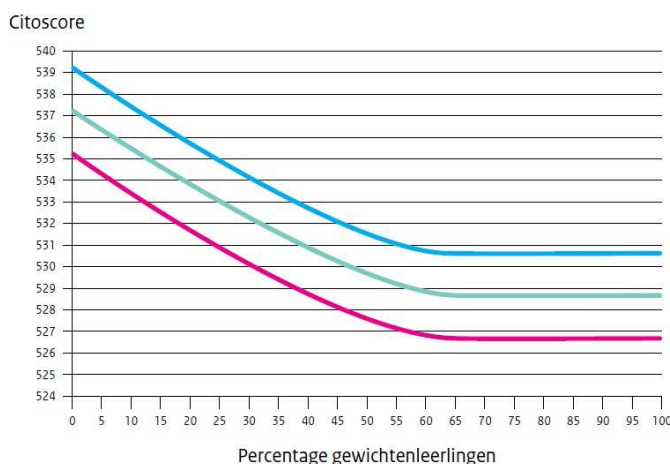
2.1 Curve met jaarlijks gelijkblijvende ondergrenzen

De inspectie gebruikt met ingang van het schooljaar 2011/2012 een nieuwe systematiek voor het beoordelen van de opbrengsten aan het eind van het basisonderwijs (OCW, 2012b). Deze systematiek, die geldt voor scholen die gebruik maken van de Cito Eindtoets, is inzichtelijker voor scholen dan de vorige en leidt niet langer tot jaarlijks wisselende ondergrenzen van scores die scholen moeten halen. Of de gemiddelde ongecorrigeerde schoolscore op de Eindtoets voldoende is, kan een school nu eenvoudig zelf bepalen door het percentage gewichtenleerlingen als uitgangspunt te nemen en vervolgens te bekijken welke gemiddelde schoolscore daar minimaal bij hoort (zie bijlage 2). Bij ieder percentage is een boven- en een ondergrens gegeven (grafiek 1). De inspectie beoordeelt de opbrengsten in een schooljaar als voldoende, als de school op of boven de ondergrens scoort.

Grafiek 1

De ondergrens, bovengrens en het landelijk gemiddelde bij verschillende percentages gewichtenleerlingen

- bovengrens
- landelijk gemiddelde
- ondergrens



Bron: Inspectie van het Onderwijs, 2011

De curve in grafiek 1 is als volgt tot stand gekomen:

- Op basis van de schoolscores van drie opeenvolgende schooljaren van ruim 6.000 basisscholen heeft de inspectie berekend wat de gemiddelde schoolscore op de Eindtoets Basisonderwijs is bij bepaalde percentages gewichtenleerlingen.

- Als door al die punten een lijn wordt getrokken die de landelijk gemiddelde schoolscore per percentage gewichtenleerlingen representeert, ontstaat een curve of een 'kromme': de groene lijn in de grafiek. Van de 'kromme' zijn vervolgens onder- en bovengrenzen afgeleid die niet meer jaarlijks aangepast hoeven te worden omdat Cito de Eindtoets longitudinaal ikt (zodat resultaten van opeenvolgende jaren aan elkaar kunnen worden gerelateerd).
- De grenswaarde van de standaardscore waaronder scholen onvoldoende presteren, ligt 2 punten onder het gemiddelde van de schoolgroep. Deze grens is zo bepaald dat scholen in ieder geval niet strenger beoordeeld worden dan voorheen. Dat wil zeggen dat de nieuwe systematiek leidt tot vergelijkbare aantallen scholen met onvoldoende eindopbrengsten als in de oude beoordelingssystematiek.

Geen onderscheid gewichten 0,3 en 1,2

Bij het berekenen van de gemiddelde schoolscore bij bepaalde percentages gewichtenleerlingen heeft de inspectie zich gebaseerd op gegevens van DUO (leerlinggewichten zoals opgegeven door scholen op de jaarlijkse teldata van 1 oktober). Er is daarbij geen onderscheid gemaakt tussen 1,2 leerlingen en 0,3 leerlingen, om de volgende redenen:

- De landelijk gemiddelde scores op de Eindtoets groeien voor 0,3- en 1,2-leerlingen steeds meer naar elkaar toe, omdat 1,2-leerlingen in verhouding tot 0,3-leerlingen steeds beter presteren. Het verschil tussen beide groepen leerlingen is bij de Eindtoets 2012 teruggelopen tot 0,7 punt. Gemiddeld halen leerlingen met gewicht 0,3 een score van 529,6 en leerlingen met gewicht 1,2 een score van 528,9 (Cito, 2012). In 2010 bedroeg het verschil tussen beide groepen nog 1,8 punt (Cito, 2011).
- Het percentage gewichtenleerlingen is voor scholen een inzichtelijk kengetal, dat zij eenvoudig zelf kunnen berekenen en kunnen betrekken bij opbrengstgericht werken. Als in die berekening een weging moet worden verdisconteerd voor 0,3- en 1,2-leerlingen, wordt het kengetal minder inzichtelijk en is de berekening minder eenvoudig door scholen zelf uit te voeren.

Percentage gewogen leerlingen op schoolniveau versus niveau groep 8

De huidige inspectiesystematiek maakt gebruik van het percentage gewogen leerlingen op de school als geheel, niet van het percentage gewogen leerlingen in groep 8.

- Een praktische reden hiervoor is, dat de inspectie niet beschikt over dat precieze percentage gewogen leerlingen in groep 8. De onderwijsnummerbestanden zijn op dit punt nog onvoldoende betrouwbaar gevuld.
- Een meer inhoudelijke reden is, dat een systematiek op het niveau van groep 8 tot grotere meetfouten leidt door het kleinere aantal leerlingen.

De inspectie heeft wel exploratief onderzocht of correctie voor gewogen leerlingen op schoolniveau tot andere beoordelingen leidt dan correctie voor gewogen leerlingen op groepsniveau. Voor het percentage gewogen leerlingen op groepsniveau is gebruik gemaakt van gegevens over de groep 11- tot 14-jarige leerlingen per school. Als op basis daarvan een curve geschat wordt, zoals dat in grafiek 1 gebeurd is voor het schoolniveau, leidt dat tot een half procent meer scholen met een onvoldoende beoordeling van de opbrengsten. In totaal zou 1,5 tot 2 procent van de scholen een andere beoordeling van de opbrengsten krijgen door over te gaan op een correctie op groepsniveau.

2.2 Onderscheid risicoanalyse en beoordeling van scholen

Als scholen onvoldoende opbrengsten bij hun leerlingen bereiken, betekent dit niet automatisch dat de inspectie deze scholen zwak of zeer zwak noemt. De inspectie gebruikt opbrengstgegevens om risico's te bepalen. Jaarlijks bespreekt de inspectie de scholen die volgens de risicoanalyse onvoldoende resultaten laten zien met de besturen die voor deze scholen verantwoordelijk zijn. Daarna kan het nodig zijn onderzoek op de school uit te voeren.

- Als scholen eenmaal onvoldoende presteren, attendeert de inspectie besturen daarop.
- Presteert een school twee keer onvoldoende, dan krijgt het bestuur een formele waarschuwing van de inspectie.
- Van een *beoordeling* 'onvoldoende' is dan nog geen sprake, omdat die beoordeling pas in beeld komt als scholen er niet in slagen om in drie jaar tijd minimaal één keer boven de ondergrens te presteren.

Scholen die drie jaar achtereen onder de voor hen geldende ondergrens presteren, worden nader door de inspectie onderzocht. In een dergelijk onderzoek kijkt de inspectie niet alleen naar opbrengsten, maar ook naar het onderwijsleerproces op de school. Het is mogelijk dat scholen valide redenen kunnen aanvoeren voor de onvoldoende prestaties, zoals een bijzondere leerlingbevolking. Ook kan herberekening aan de orde zijn, als leerlingen hebben meegedaan aan de Eindtoets die de inspectie niet meetelt in de gemiddelde score (zoals leerlingen die naar het praktijkonderwijs uitstromen). Als herberekenen van de opbrengsten niet aan de orde is, blijven deze onvoldoende en wordt de school, zolang hooguit één andere normindicator onvoldoende is, zwak. Zijn twee of meer normindicatoren onvoldoende naast de opbrengsten, dan wordt de school zeer zwak.

Scholen krijgen dus nooit het oordeel zwak of zeer zwak enkel en alleen op basis van de risicoanalyse. Onvoldoende opbrengsten zoals deze uit de risicoanalyse naar voren komen, leiden nooit automatisch tot geïntensiveerd toezicht. Informatie van het bestuur en informatie die de inspectie zonodig door middel van een schoolbezoek vergaart, worden hierbij betrokken.

3 Onderzoeksopzet

Voor de secundaire analyses heeft de inspectie gebruik gemaakt van eigen gegevens die gekoppeld zijn aan data uit het COOL-cohortonderzoek (3.1). De gecombineerde gegevens zijn gebruikt om de effecten van verschillende modellen voor correctiefactoren te onderzoeken. Deze modellen zijn steeds afgezet tegen de huidige werkwijze van de inspectie (3.2). Nagegaan is hoeveel variantie de modellen verklaren in de opbrengstverschillen tussen scholen. Ook is onderzocht hoeveel scholen een andere beoordeling krijgen en welke scholen dit zijn. Vervolgens is nagegaan in hoeverre de onderzochte modellen voldoen aan enkele criteria die voor het beoordelen van opbrengsten belangrijk zijn (3.3).

3.1 Beschikbare gegevens voor de secundaire analyses

Inspectiegegevens: prestaties, leerlinggewicht en etniciteit op schoolniveau

De inspectie beschikt over de gemiddelde schoolscores op de Eindtoets in groep 8. Deze scores krijgt de inspectie jaarlijks van alle scholen die de Eindtoets gebruiken. Daarnaast weet de inspectie via DUO hoeveel gewogen leerlingen een school heeft. Aan de hand van deze twee gegevens wordt nu jaarlijks door middel van een risicoanalyse nagegaan of een school onvoldoende dan wel voldoende presteert (zie vorige hoofdstuk). De etniciteit van ouders maakt geen deel meer uit van de gewichtenregeling, maar informatie hierover zit wel in de onderwijsnummerbestanden die de inspectie heeft. Deze gegevens zijn naar het schoolniveau geaggregeerd. Van iedere school is daardoor bekend hoeveel procent autochtonen in de leerlingbevolking zijn vertegenwoordigd en hoeveel procent westerse en niet-westerse allochtonen.

COOL-data: opleiding ouders

Om aan meer gegevens over de opleidingsniveaus van ouders te komen is gebruik gemaakt van data uit het COOL-cohort (Driessen, Mulder en Roeleveld, 2012). Via de COOL-onderzoekers kreeg de inspectie de beschikking over vier percentages: de percentages leerlingen per school van wie de ouders hooguit lager onderwijs hebben, een opleiding op lbo-niveau, een opleiding op mbo-niveau of een opleiding op het niveau van het hoger onderwijs. Deze gegevens dateren uit 2011 en zijn op basis van de leerlingenadministraties door basisscholen aan de COOL-onderzoekers verstrekt (gegevens van ouders zelf, verzameld via COOL-oudervragenlijsten, bevatten teveel ontbrekende data om te gebruiken). De percentages zijn berekend op basis van de leerlingen in de groepen 2, 5 en 8 en dus niet op basis van alle leerlingen van de school. Ze worden hier beschouwd als betrouwbare benaderingen van de percentages die voor de gehele school zouden gelden: de correlatie tussen het percentage gewogen leerlingen per school en het percentage gewogen leerlingen in de groepen 2, 5 en 8 is hoog, namelijk .90. Het inkomen van ouders is niet als aparte correctiefactor meegenomen, omdat recente analyses uitwijzen dat opleidingsniveau en etnische herkomst betere voorspellers zijn (Roeleveld, Mooij en Fettelhaar, 2011).

Onderzochte steekproef van scholen

Het scholenbestand van COOL is het uitgangspunt voor de secundaire analyses. Dit bestand bevat 553 scholen. Na een selectie op scholen met tenminste 20 leerlingen en op scholen waar van tenminste 75 procent van de gezinnen de hoogste opleiding bekend is, kreeg de inspectie een bestand van 514 scholen. Van deze scholen moeten, om tot een oordeel over de opbrengsten in groep 8 te kunnen komen, ook

gegevens beschikbaar zijn over de gemiddelde Eindtoetsscores in 2011, 2010 en 2009. Dat is voor 387 scholen het geval. Voor de analyses komen verder alleen scholen in aanmerking die vanuit groep 8 jaarlijks 10 of meer leerlingen laten uitstromen naar het voortgezet onderwijs. De onderzochte groep bestaat na toepassing van dit criterium uiteindelijk nog uit 347 scholen.

Deze groep scholen is niet zonder meer representatief voor de Nederlandse populatie van basisscholen. Ten eerste zijn alle beschikbare COOL-scholen gebruikt, zowel de representatieve steekproef die COOL onderscheidt als de aanvullende steekproef van scholen die veel achterstandsleerlingen hebben. Ten tweede blijven scholen die de Eindtoets niet afnemen buiten beschouwing, evenals kleine scholen. De onderzochte groep bestaat voor ongeveer twee derde uit scholen die meer dan 10 procent gewogen leerlingen hebben, maar landelijk heeft slechts 41 procent van de scholen meer dan 10 procent gewogen leerlingen. Door deze enigszins scheve verdeling kunnen extrapolaties van aantallen scholen met een onvoldoende beoordeling van de opbrengsten vanuit de steekproef naar landelijk niveau slechts met de nodige reserves worden gemaakt. Relaties tussen correctiefactoren kunnen daarentegen wel gelegd worden, net zoals vergelijkingen tussen effecten van modellen gemaakt kunnen worden.

3.2 Onderzochte modellen voor het corrigeren van opbrengsten

Beschikbaar voor de analyses zijn, naast de gemiddelde scores van scholen op de Eindtoets in 2011, 2010 en 2009, de volgende gegevens op schoolniveau:

- Leerlinggewicht (percentage gewogen leerlingen),
- Etniciteit (percentage westerse en niet-westerse allochtonen),
- Opleidingsniveau (vier percentages per school: ouders met maximaal opleidingsniveau lager onderwijs, lager beroepsonderwijs, middelbaar beroepsonderwijs en hoger onderwijs).

Gegevens over andere factoren die mogelijk ook relevant zijn voor prestaties, zoals de intelligentie of de motivatie van leerlingen of specifieke problematische omstandigheden in gezinnen, zijn niet beschikbaar en blijven dus ook in alle onderzochte modellen buiten beschouwing.

Op basis van de beschikbare gegevens zijn zeven verschillende modellen van correctiefactoren onderzocht en afgezet tegen de huidige systematiek van de inspectie (tabel 2). Al deze modellen beperken zich tot kenmerken op schoolniveau.

Tabel 2 Onderzochte modellen van correctiefactoren

Model	Omschrijving
0	Leerlinggewicht (= huidige systematiek)
1	Leerlinggewicht en etniciteit
2	Opleidingsniveau
3	Opleidingsniveau en etniciteit
4	Leerlinggewicht, etniciteit en opleidingsniveau
5	Percentage hoger opgeleiden (variant opleidingsniveau: spiegel van de huidige

	gewichtenregeling)
6	Zeven schoolgroepen (gebaseerd op Roeleveld, Mooij en Fettelaar, 2011): • meer dan 50 % ouders met maximaal lbo én 50 % of meer allochtoon • meer dan 50 % ouders met maximaal lbo én meer dan 50 % autochtoon • meer dan 75 % maximaal lbo of mbo, maar geen van beide meer dan 50 % • meer dan 50 % ouders met maximaal mbo • meer dan 75 % maximaal mbo of ho, maar geen van beide meer dan 50 % • meer dan 50 % ouders met ho • gemengd, geen dominante groep
7	Combinatiemodel: scholen met minder dan 10 % ongewogen leerlingen zijn volgens de huidige inspectiesystematiek beoordeeld, maar de groep scholen met meer dan 90 % ongewogen leerlingen is opgesplitst in vier subgroepen afhankelijk van het percentage hoger opgeleide ouders: • a. minder dan 32 procent hoogopgeleide ouders, • b. tussen 32 en 48 procent hoogopgeleide ouders, • c. tussen 48 en 64 procent hoogopgeleide ouders • d. meer dan 64 procent hoogopgeleide ouders.

Bij de modellen 1, 2 en 3 is gebruik gemaakt van wisselende variabelen, bij model 4 zijn alle beschikbare variabelen tegelijk onderzocht. Model 5 is gebaseerd op het percentage hoger opgeleide ouders op een school; dit model is daardoor een spiegel van de huidige gewichtenregeling, die uitgaat van het percentage lager opgeleide ouders. Model 6 gaat uit van een indeling van scholen in schoolgroepen, die gebaseerd is op de vertegenwoordiging van verschillende groepen leerlingen in de school. Bij dit model zijn opleidingsgegevens en gegevens over etniciteit gebruikt.

Bij de genoemde modellen is de ondergrens 2 punten onder het gemiddelde van vergelijkbare scholen gelegd. Net zoals in de huidige inspectiesystematiek, waar scholen met een gelijk percentage gewogen leerlingen eenzelfde ondergrens hebben, is in elk onderzocht model de ondergrens voor scholen met gelijke kenmerken steeds hetzelfde. Een score onder de ondergrens leidt tot een onvoldoende beoordeling in een bepaald schooljaar, een score gelijk aan of boven de ondergrens leidt tot een voldoende beoordeling in een bepaald schooljaar. Van een onvoldoende beoordeling van de opbrengsten van een school is uiteindelijk pas sprake als een school in alle drie onderzochte schooljaren onder de ondergrens presteert.

Bij model 7 zijn de scholen eerst verdeeld in twee groepen: scholen met minder dan 90 procent ongewogen leerlingen en scholen met 90 procent of meer ongewogen leerlingen. De eerste groep is beoordeeld volgens de huidige inspectiesystematiek. In de tweede groep is gekeken naar het percentage hoogopgeleide ouders, dat hier gemiddeld genomen op 48 ligt. Deze tweede groep is vervolgens in vier subgroepen verdeeld (zie tabel 2 voor de specificatie van subgroepen). Per subgroep is de gemiddelde score op de Eindtoets berekend; de ondergrens per subgroep is bepaald door twee punten van dit gemiddelde af te trekken. Model 7 is geïntroduceerd om in de groep scholen met 90 procent of meer ongewogen leerlingen meer differentiatie aan te brengen. De reden hiervoor is dat soms wordt gesteld dat scholen met veel lager opgeleide ouders relatief vaak een oordeel onvoldoende op de opbrengsten krijgen. Dit model zou duidelijk moeten maken of dat het geval is.

3.3 Analyses

In de analyses is ten eerste nagegaan hoeveel variantie in prestaties van scholen verklaard wordt door de diverse modellen. Hoe hoger het percentage verklaarde variantie, hoe beter de gebruikte factoren corrigeren en hoe beter de feitelijke bijdrage van scholen vervolgens in beeld komt.

Ten tweede is voor ieder model nagegaan hoeveel verschillen in beoordelingen van scholen optreden in vergelijking met de huidige inspectiesystematiek. Dat wordt duidelijk door per model te inventariseren hoeveel scholen een voldoende of onvoldoende beoordeling krijgen van hun opbrengsten over drie opeenvolgende schooljaren. Voor alle modellen geldt dat een school pas een onvoldoende beoordeling van de opbrengsten krijgt, als de prestaties in alle drie schooljaren onvoldoende zijn.

Ten derde is nagegaan in hoeverre de modellen voldoen aan eisen die aan correctiefactoren gesteld kunnen worden (zie ook Dijkstra, Karsten, Veenstra en Visscher, 2001):

Verklarende kracht/relevantie: de factoren moeten de bijdrage van scholen aan prestaties beter kunnen schatten dan de huidige inspectiesystematiek,

Eerlijkheid: de factoren moeten de bijdrage van scholen aan opbrengsten eerlijk en fair benaderen,

Dekking: de factoren moeten voor de hele populatie van leerlingen en basisscholen beschikbaar zijn,

Betrouwbaarheid: de factoren moeten betrouwbaar gemeten kunnen worden (geen grote administratieve fouten of meetproblemen),

Praktische haalbaarheid: de factoren moeten zonder al te veel belasting van scholen kunnen worden bepaald,

Transparantie: de berekening van factoren moet inzichtelijk zijn voor scholen en door scholen zelf gereproduceerd kunnen worden.

Het criterium van transparantie is vooral van belang om scholen te ondersteunen bij opbrengstgericht werken. Het is nodig dat scholen ruim van tevoren weten wat de minimumeis is voor hun opbrengsten, maar ook is het nodig dat ze kunnen zien wat de gemiddelde en bovengemiddelde opbrengsten van scholen met een vergelijkbare leerlingbevolking zijn. Deze gegevens kunnen besturen en scholen gebruiken om te bepalen welke resultaten of doelen ze willen bereiken.

4 Resultaten van de secundaire analyses

Dit hoofdstuk beschrijft eerst wat de analyses hebben opgeleverd in termen van verklaarde variantie van de onderzochte modellen (4.1) en verschillen in de beoordeling van scholen (4.2). Daarna worden de resultaten per model afgezet tegen de criteria die gesteld kunnen worden, zoals transparantie voor scholen (4.3).

4.1 Verklaarde variantie

Het percentage verklaarde variantie per model geeft aan hoe sterk de correctie op grond van bepaalde factoren in dat specifieke model is. Dit percentage laat zien hoeveel procent van de opbrengstverschillen tussen scholen toegeschreven kan worden aan de gebruikte correctiefactoren. Hoe hoger het percentage verklaarde variantie, hoe beter de correctie is voor de niet-schoolse factoren en hoe beter dus de feitelijke bijdrage van scholen aan opbrengsten van leerlingen in beeld komt.

De modellen leiden tot verschillende percentages verklaarde variantie (tabel 3).

Tabel 3 Verklaarde variantie per model

	Model	Percentage verklaarde variantie tussen scholen
0	Leerlinggewicht (huidige systematiek)	48
1	Leerlinggewicht en etniciteit	49
2	Opleidingsniveau	51
3	Opleidingsniveau en etniciteit	54
4	Leerlinggewicht, etniciteit en opleidingsniveau	55
5	Percentage hoger opgeleiden	41
6	7 schoolgroepen (dummy's gebruikt)	44
7	Combinatiemodel	Variabel, maar maximaal 50

De huidige inspectiesystematiek leidt tot 48 procent door de correctiefactoren verklaarde variantie. Dat betekent dat prestatieverschillen tussen scholen voor bijna de helft toe te schrijven zijn aan het percentage gewogen leerlingen. De andere helft is toe te schrijven aan verschillen in de kwaliteit van het onderwijs en eventuele andere factoren waar niet voor gecorrigeerd kon worden, zoals bijvoorbeeld de motivatie van leerlingen en buitenschoolse factoren.

Twee modellen zitten onder de verklaarde variantie van de huidige inspectiesystematiek (modellen 5 en 6), hetgeen betekent dat zij niet leiden tot een betere schatting van de bijdrage van scholen aan prestaties. De andere modellen (1 tot en met 4) resulteren alle in hogere percentages verklaarde variantie dan de huidige inspectiesystematiek.

Bij model 7 is de verklaarde variantie bepaald door dummy variabelen te maken van het percentage gewichtenleerlingen per school (meer dan 10 procent, minder dan 10 procent) en deze in een regressievergelijking op te nemen en te kruisen met ofwel het percentage gewogen leerlingen, ofwel het percentage hoger opgeleiden. Bij model 7 varieert de verklaarde variantie als de grens voor het onderscheid tussen beoordeling op basis van het percentage gewogen leerlingen en het percentage

hoger opgeleide ouders anders wordt gelegd. Wanneer deze grens dichterbij 100 procent ongewogen leerlingen ligt, stijgt de verklaarde variantie. De verklaarde variantie komt dan ook rond de 50 procent te liggen. Het aantal scholen in de steekproef met een hoog percentage ongewogen leerlingen is echter dermate laag, dat voor verdere analyses is gekozen deze grens te handhaven op 90 procent.

Toevoegen van etniciteit aan de huidige inspectiesystematiek leidt, zo blijkt uit tabel 3, tot 1 procent meer verklaarde variantie. Het opleidingsniveau verklaart als correctiefactor meer variantie dan het leerlinggewicht en het leerlinggewicht samen met etniciteit. Dat ligt voor de hand, omdat het opleidingsniveau meer gedifferentieerde informatie over de ouders van een school bevat dan het leerlinggewicht. De uitschieter is model 4 met 55 procent verklaarde variantie. Dat juist dit model de bijdrage van scholen het beste schat ligt eveneens voor de hand, omdat hier alle drie soorten beschikbare gegevens (leerlinggewicht, etniciteit en opleidingsniveau) zijn benut. Dit resultaat komt overeen met de bevindingen van Roeleveld, Mooij, Fettelaar en Ledoux (2011).

4.2 Verschillen in beoordeling van scholen

Aantallen en percentages onvoldoende beoordeelde scholen

Per model is nagegaan in hoeverre de beoordeling van opbrengsten verschillend uitpakt in vergelijking met de huidige inspectiesystematiek, die leidt tot 7,2 procent onvoldoende scholen in de COOL-steekproef (tabel 4).

Tabel 4 Oordeel over de opbrengsten van scholen (n=347) op basis van de huidige inspectiesystematiek (leerlinggewicht)

% scholen onvoldoende in 2011	26
% scholen onvoldoende in 2010	27
% scholen onvoldoende in 2009	32
% onvoldoende in alle drie jaren	7,2

Tabel 4 laat zien dat bij de huidige inspectiesystematiek jaarlijks een kwart tot een derde van de COOL-scholen onder de ondergrenzen scoort. De groep scholen die in alle drie jaren onder de ondergrens scoort is veel kleiner, namelijk 7,2 procent. Dit zijn de scholen die op basis van de risicoanalyse in aanmerking komen voor verder onderzoek door de inspectie. Op basis van dat verdere onderzoek kan de inspectie besluiten scholen als zwak of zeer zwak te beoordelen en het toezicht te intensiveren (zie 2.2).

De verschillende modellen leiden tot verschillende percentages onvoldoende scholen ten opzichte van de huidige werkwijze (tabel 5).

Tabel 5 Oordeel over de opbrengsten van scholen (n=347) over drie schooljaren in de verschillende modellen

	0	1	2	3	4	5	6	7
% scholen onvoldoende in 2011	26	25	24	24	24	27	25	24
% scholen onvoldoende in 2010	27	28	30	30	30	30	29	28
% scholen onvoldoende in 2009	32	32	31	29	28	31	31	31
% scholen onvoldoende in alle drie jaren	7,2	6,9	6,9	7,2	6,6	8,1	8,2	6,3

Er zit wat fluctuatie in de percentages onvoldoende scorende scholen per jaar in de verschillende modellen en dat geldt ook voor de percentages onvoldoende scorende scholen in alle drie jaren. Het verschil met de huidige werkwijze is nooit groter dan 1 procent. De modellen die minder variantie verklaren dan de huidige systematiek (5 en 6) leiden tot iets meer onvoldoendes bij de scholen. De modellen die meer variantie verklaren dan de huidige systematiek (1 tot en met 4) leiden tot een iets lager of even hoog percentage onvoldoendes bij scholen. Model 7 geeft het laagste percentage scholen met onvoldoendes.

Model 4, dat de meeste variantie verklaart, resulteert in 6,6 procent scholen met onvoldoende opbrengsten over drie jaar. Het verschil met de huidige inspectiesystematiek is 0,6 procent. Op de steekproef van 347 COOL-scholen zijn volgens de huidige systematiek 25 scholen als onvoldoende te klassificeren, volgens model 4 zou het gaan om 23 scholen.

Aantallen en percentages scholen die van beoordeling veranderen

Het percentage onvoldoende scholen zegt nog niet alles, omdat het bij eenzelfde percentage onvoldoende scholen in verschillende modellen om verschillende scholen kan gaan. Met andere woorden: in de huidige systematiek zijn 25 van de COOL-scholen onvoldoende, in model 4 zijn het er bijna evenveel maar het kunnen wel (deels) andere scholen zijn. Daarom is nagegaan hoeveel scholen per model van voldoende naar onvoldoende verschuiven en andersom (tabel 6).

Omdat de modellen 5 en 6 minder verklaarde variantie laten zien dan het huidige model, zijn zij niet in onderstaande analyse meegenomen.

Tabel 6 Verschuivingen van scholen (n=347) in verschillende modellen ten opzichte van de huidige inspectiesystematiek

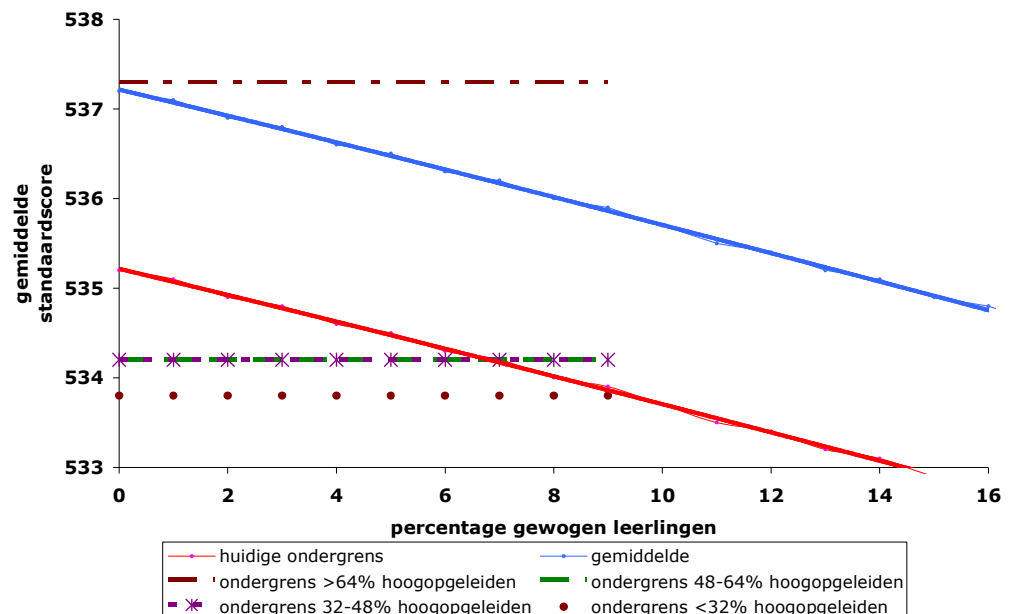
Model	1	2	3	4	5	6	7
Aantal scholen van onvoldoende naar voldoende	3	7	5	6			2
Aantal scholen van voldoende naar onvoldoende	2	6	5	4			0
Totaal aantal verschuivende scholen	5	13	10	10			2
Percentage verschuivende scholen	1	4	3	3			<1

De modellen verschillen in het aantal scholen dat van onvoldoende naar voldoende gaat en ook in het aantal scholen dat van voldoende naar onvoldoende gaat. In percentages uitgedrukt verschuift minder dan 1 tot maximaal 4 procent van de steekproef. Omdat scholen met achterstandsleerlingen in de COOL-steekproef oververtegenwoordigd zijn en omdat een deel van de steekproef buiten de analyses valt omdat er geen Eindtoetsgegevens zijn of omdat er te weinig leerlingen in groep 8 zitten, is het niet goed mogelijk om deze percentages betrouwbaar naar de populatie te extrapoleren. Om toch een indruk te geven van de aantallen scholen waar het in de populatie om zou kunnen gaan: 1 procent staat gelijk aan ongeveer 70 basisscholen.

In grafiek 2 staan de ondergrenzen weergegeven voor de vier naar percentage hoger opgeleiden onderscheiden groepen scholen in model 7. De ondergrenzen van de groepen 32-48 procent hoger opgeleiden per school en 48-64 procent hoger opgeleiden vallen samen en liggen 0,3 punt boven de ondergrens van de groep met minder dan 32 procent hoger opgeleiden. De ondergrens van de groep scholen met meer dan 64 procent hoger opgeleiden ligt boven de landelijk gemiddelde

schoolscore voor scholen die geen gewogen leerlingen hebben. De verschuivingen van scholen in model 7 zijn zeer beperkt: er gaan geen scholen van voldoende naar onvoldoende en slechts twee scholen verschuiven van onvoldoende naar voldoende. Deze twee scholen bestaan beide voor meer dan de helft uit leerlingen van hoger opgeleide ouders.

Grafiek 2 Ondergrenzen voor voldoende in model 7



Dat er geen scholen uit de groep met minder dan 32 procent hoger opgeleide ouders verschuiven van onvoldoende naar voldoende is een opvallend resultaat, omdat soms verondersteld wordt dat dit juist het geval zou zijn bij een betere correctie voor het opleidingsniveau van de ouders.

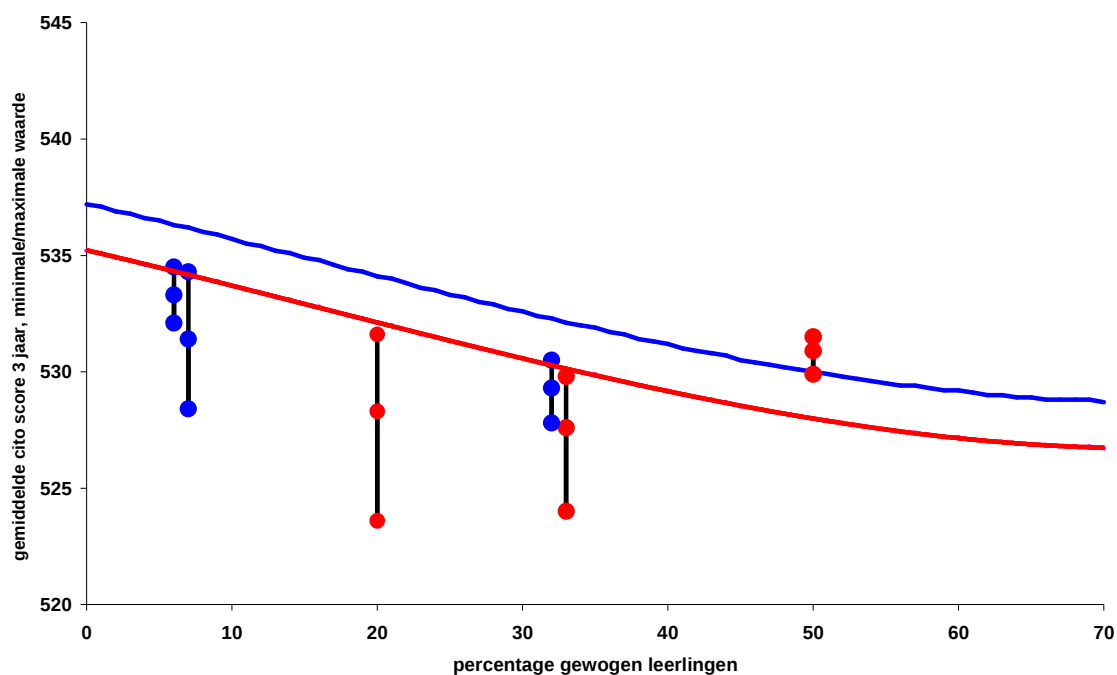
Welke scholen krijgen een andere beoordeling?

Over alle modellen heen bekeken zijn 315 van de 347 scholen steeds als voldoende beoordeeld en 17 scholen steeds als onvoldoende. In totaal worden dus 332 scholen ofwel 96 procent van de COOL-steekproef consistent van dezelfde beoordeling voorzien. Van de steekproef verschuift 4 procent ofwel 13 scholen van oordeel in een of meer van de modellen. Bij 6 van deze 13 scholen gebeurt dat twee of drie keer. Deze 6 scholen zijn dus het meest gevoelig voor het model van opbrengstenbeoordeling.

In grafiek 3 is te zien om welke 6 scholen het gaat. Elke staaf in de grafiek staat voor één school. De top van de staaf geeft de hoogste score weer die die specifieke school in de afgelopen 3 jaar heeft behaald, de onderkant van de staaf geeft de laagste score weer. Het bolletje op de staaf is de gemiddelde score van deze scholen, over 3 jaar berekend.

Grafiek 3

Samenstelling leerlingbevolking in percentage ongewogen leerlingen (X-as) en gemiddelde score op de Eindtoets over 3 schooljaren (Y-as) van de 6 scholen die vaak van beoordeling wisselen afhankelijk van het gebruikte model



Rood= 3 keer onvoldoende, 2 keer voldoende

Blauw= 2 keer onvoldoende, 3 keer voldoende

De wisselend beoordeelde groep scholen is geen homogene groep, maar loopt zowel qua leerlingbevolking als prestaties uiteen. Er zitten geen scholen met een meerderheid aan gewogen leerlingen bij. Het is dus niet deze groep scholen die snel anders geclassificeerd wordt bij een ander model voor opbrengstenbeoordeling. Het zijn echter ook niet alleen de scholen met veel ongewogen leerlingen, ofwel de scholen waar meer hoogopgeleide ouders verwacht kunnen worden.

Scholen die in het sterkst corrigerende model van oordeel veranderen

Om een nog concreter beeld te krijgen van de scholen die van beoordeling veranderen, is ingezoomd op model 4, omdat dit model het beste de bijdrage van de school aan prestaties van leerlingen schat. In model 4 verschuiven 10 scholen van beoordeling: 6 scholen gaan van een onvoldoende naar een voldoende en 4 scholen van een voldoende naar een onvoldoende. Tabel 7 geeft een beeld van de 6 scholen die van onvoldoende naar voldoende gaan.

Tabel 7 Kenmerken van 6 scholen die door model 4 van onvoldoende naar voldoende gaan

	Leerlinggewicht			Opleiding				Etniciteit
	% 0,0	% 0,3	% 1,2	% lo	% lbo	% mbo	% ho	% niet-westers
A	21	4	75	51	14	25	10	93
B	41	9	50	31	29	32	8	86
C	66	13	21	19	20	45	16	67
D	75	11	14	14	14	48	24	66
E	80	9	11	6	16	56	22	40
F	97	3	0	0	4	45	51	6
Steek-proef	78	11	11	8	16	42	34	26

Zoals eerder gezegd is de veronderstelling vaak dat vooral scholen met veel leerlingen die laag opgeleide ouders hebben ten onrechte als onvoldoende worden aangemerkt. Van deze 6 scholen voldoen school A en B redelijk aan dat beeld. Deze scholen hebben (in vergelijking met het gemiddelde in de steekproef) zeer veel leerlingen met laagopgeleide ouders en weinig leerlingen met mbo- of hoger opgeleide ouders. Voor C, D en E is dat minder het geval. Een overeenkomst tussen de vijf scholen is wel het relatief hoge percentage niet-westerse allochtone leerlingen in de leerlingbevolking, dat niet meer terug te zien is in de gewichtenregeling omdat etniciteit niet meer medebepalend is voor het gewicht. School F laat een heel ander beeld zien dan de andere scholen. Deze school heeft nauwelijks gewogen leerlingen of niet-westerse allochtonen en de meerderheid van de ouders is hoogopgeleid.

In discussies over de opbrengstenbeoordeling van de inspectie is verder vaak de verwachting dat vooral scholen met veel hoogopgeleide ouders door een betere schatting van de bijdrage van scholen aan prestaties van beoordeling zouden wisselen, dat wil zeggen vaker een onvoldoende beoordeling zouden krijgen. Tabel 8 geeft een beeld van de 4 scholen die van voldoende naar onvoldoende gaan.

Tabel 8 Kenmerken van 4 scholen die door model 4 van voldoende naar onvoldoende gaan

	Leerlinggewicht			Opleiding				Etniciteit
	% 0,0	% 0,3	% 1,2	% lo	% lbo	% mbo	% ho	% niet-westers
G	68	19	13	11	16	57	16	26
H	94	5	1	1	7	36	56	8
I	93	6	1	0	9	38	53	6
J	63	35	2	2	35	40	23	3
Steek-proef	78	11	11	8	16	42	34	26

Van de vier scholen voldoen H en I redelijk aan de geschetste verwachting: beide scholen hebben in meerderheid leerlingen met hoogopgeleide ouders. Op de scholen G en J is dat echter niet het geval. Een overeenkomst tussen H, I en J is nog wel het zeer geringe percentage niet-westerse allochtone leerlingen. Evenals bij de vorige groep (zie tabel 8) lijkt echter ook hier de factor etniciteit van belang. De meeste verschuivende scholen hebben in vergelijking met het steekproefgemiddelde zeer weinig niet-westerse allochtone leerlingen.

Inzoomen op de scholen die van beoordeling veranderen in de verschillende modellen maakt vooral uit dat dit een heterogene groep is. Het is dus niet zo dat vooral scholen met veel gewogen leerlingen van een onvoldoende naar een voldoende gaan en het is ook niet zo dat vooral scholen met veel hoogopgeleide ouders van een voldoende naar een onvoldoende gaan.

4.3 Eisen aan correctiefactoren en modellen

In hoofdstuk 3 is geschetst aan welke eisen correctiefactoren moeten voldoen. De onderzochte modellen kunnen nu getoetst worden aan deze eisen (tabel 9).

De eerste eis is dat de modellen een grotere verklarende kracht moeten hebben, en dus beter moeten corrigeren, dan de huidige systematiek. Voor de modellen 5 en 6 geldt dat niet. Deze modellen vallen alleen om die reden al af. Voor de overige modellen zijn de resterende eisen nog relevant. Deze modellen voldoen alle aan de eisen van eerlijkheid, dekking en betrouwbaarheid. De modellen leiden alle tot een zorgvuldige schatting van de bijdrage van scholen aan prestaties (eerlijkheid) en de benodigde gegevens kunnen in principe bij alle scholen opgevraagd worden (dekking).

De modellen zijn verder gelijk op het punt van de betrouwbaarheid; scholen kunnen fouten maken bij het aanleveren van gegevens over de opleiding van ouders en de etniciteit, maar dat gebeurt ook met de gegevens die de gewichtenregeling vereist (Onderwijsverslag, 2012). De gegevens die nodig zijn voor alle modellen zouden opgevraagd kunnen worden via een procedure die vergelijkbaar is met die van de huidige gewichtenregeling en die een formele status heeft (praktische haalbaarheid).

Op het punt van de transparantie levert ieder model dat meer dan één variabele bevat problemen op, omdat het voor scholen lastiger wordt om zelf te bepalen aan welke opbrengstennorm ze moeten voldoen. De modellen 1 tot en met 4 verliezen het op dit punt van de huidige systematiek. Ook voor model 7, dat op het opleidingsniveau voortbouwt, geldt dit omdat voor elke school naast de gewichtenregeling ook bekend moet zijn hoeveel hoger opgeleide ouders er zijn.

Tabel 9 Toetsing van modellen aan eisen die aan correctiefactoren gesteld kunnen worden

Model	1	2	3	4	5	6	7
Grotere verklarende kracht dan huidige model	ja	ja	ja	ja	nee	nee	*
Eerlijkheid	ja	ja	ja	ja			ja
Dekking	ja	ja	ja	ja			ja
Betrouwbaarheid	ja	ja	ja	ja			ja
Praktische haalbaarheid	ja	ja	ja	ja			ja
Transparantie	nee	nee	nee	nee			nee

*: afhankelijk van grenswaarde

5 Conclusies en discussie

Verbetering correctiefactoren mogelijk

Meer gedifferentieerde informatie over het opleidingsniveau van ouders leidt tot een betere beoordeling van de bijdrage die scholen leveren aan de prestaties van leerlingen dan de huidige inspectiesystematiek, die alleen gebruik maakt van het leerlinggewicht. De huidige systematiek verklaart 48 procent van de opbrengstverschillen tussen scholen aan de hand van één factor, terwijl drie factoren samen, leerlinggewicht, opleidingsniveau en etniciteit, 55 procent kunnen verklaren. Ook het opleidingsniveau alleen doet het beter dan het leerlinggewicht (51 versus 48 procent verklaarde variantie). Er is dus zeker iets te winnen bij het gebruik van meer en andere correctiefactoren dan de huidige.

Deze conclusie is in lijn met het onderzoek van Roeleveld e.a. (2011). Uit dit onderzoek bleek verder ook dat analyses met een model waarin leerlinggegevens op individueel niveau zijn opgenomen, (iets) nauwkeuriger zijn dan analyses met alleen variabelen op schoolniveau: een dergelijk model leidt tot meer verklaarde variantie. Omdat de inspectie nu niet beschikt over gegevens op individueel leerlingniveau, beperken de analyses die in dit rapport beschreven zijn zich noodgedwongen tot kenmerken op schoolniveau.

Verschuivingen van beoordelingen bij heterogene groep scholen

In totaal zijn zeven modellen van correctiefactoren onderzocht. Bij deze modellen is het percentage scholen dat drie jaar lang onvoldoende opbrengsten behaalt redelijk vergelijkbaar: de range loopt van 6,3 tot 8,2 procent van alle scholen. Voor 96 procent van de scholen in de steekproef maakt het voor de beoordeling niet uit welk model gebruikt wordt: deze scholen worden in alle modellen consequent als voldoende of onvoldoende beoordeeld. Voor 4 procent van de scholen maakt het gebruikte model wel het verschil tussen een onvoldoende of een voldoende beoordeling.

In het best verklarende model krijgen 10 van de 347 (3 procent) scholen een andere beoordeling dan bij de huidige inspectiesystematiek: 6 scholen gaan van een onvoldoende naar een voldoende, 4 van een voldoende naar een onvoldoende. Landelijk zou dit betekenen dat zo'n 210 scholen (3 procent van in totaal circa 7.000 basisscholen) een andere beoordeling krijgen. Extrapoleren naar de populatie is echter niet goed mogelijk, omdat de onderzochte steekproef niet representatief is; deze schattingen moeten dus met de nodige voorzichtigheid bekeken worden.

Critici van de huidige inspectiesystematiek veronderstellen doorgaans, dat deze te streng uitpakt voor scholen met veel gewogen leerlingen en te soepel voor scholen waar veel kinderen van hoog opgeleide ouders zitten (die in de huidige systematiek niet te traceren zijn). De secundaire analyses bevestigen deze veronderstellingen niet. Verschuivingen van onvoldoende naar voldoende en omgekeerd beperken zich niet tot de beschreven twee groepen, maar betreffen een breder spectrum aan scholen.

Verlies aan transparantie

De huidige inspectiesystematiek is transparant voor scholen, omdat ze weten aan welke ondergrens van opbrengsten ze moeten voldoen bij hun specifieke percentage gewogen leerlingen. Uitbreiding van correctiefactoren leidt weliswaar tot meer verklaarde variantie en een iets betere schatting van de bijdrage van scholen aan opbrengsten, maar naarmate het aantal factoren toeneemt neemt de inzichtelijkheid

voor scholen weer af. Het wordt voor scholen lastig, zo niet onmogelijk om zelf te berekenen aan welke ondergrens ze moeten voldoen.

De inspectie gebruikt nu alleen het percentage gewogen leerlingen als correctiefactor. Het is zonder extra administratieve belasting van scholen mogelijk om ook de etniciteit van de leerlingen mee te nemen in de beoordeling, omdat deze gegevens al beschikbaar zijn via het onderwijsnummer. De verklaarde variantie stijgt dan met 1 procent. Dat weegt niet zonder meer op tegen het bovenbeschreven verlies aan transparantie.

Administratieve belasting van scholen

Een substantiële verbetering van correctiefactoren is mogelijk als de inspectie kan beschikken over meer gedifferentieerde gegevens over het opleidingsniveau van ouders dan wat nu in het leerlinggewicht verwerkt zit. Dat vergt echter wel degelijk een zwaardere administratieve belasting van scholen.

- Om goede opleidingsgegevens te verkrijgen, zouden scholen van alle leerlingen moeten registreren welk opleidingsniveau beide ouders hebben.
- Idealiter zou dit zo specifiek mogelijk moeten gebeuren, nog gedetailleerder dan nu in COOL is gebeurd. Zo is het wenselijk om het vmbo en het mbo onder te verdelen in niveaus en ook om binnen het hoger onderwijs een nader onderscheid te maken in hbo, wetenschappelijk onderwijs, gepromoveerd.
- Vervolgens zou onderzocht kunnen worden of er een continue schaal te maken is van opleidingsniveaus van beide ouders samen, die in de plaats kan komen van de vier variabelen op schoolniveau die nu in de secundaire analyses gebruikt zijn. Een continue variabele heeft het voordeel dat het opleidingsniveau voor elke school als één getal uitgedrukt kan worden, hetgeen de transparantie voor scholen weer ten goede zou komen.

Het heeft geen zin om bij scholen de huidige gegevens voor de gewichten op te vragen en daarnaast nog een meer gedetailleerde registratie van de opleidingen te vragen, want naast opleiding voegt gewicht nauwelijks nog iets toe. Een meer gedetailleerde bevraging zou dus in de plaats moeten komen van de huidige registratie van gegevens voor de gewichtenbepaling. Daarbij hoeft de financiering op basis van de huidige gewichten niet te veranderen; het gaat hier alleen om het verzamelen van aanvullende gegevens waaruit ook de nu gebruikte gewichten gereconstrueerd kunnen worden. Gegevens moeten in ieder geval altijd opgevraagd worden binnen een formele regeling, waardoor ook van sancties sprake kan zijn als scholen foute gegevens verstrekken.

Verplichte eindtoets

Bovenstaande exercities hebben alleen zin als er ook een verplichte eindtoets voor alle basisscholen komt. Nu vallen sommige scholen buiten de beoordelingssystematiek zoals die voor de overgrote meerderheid van scholen geldt: 15 procent gebruikt de Eindtoets Basisonderwijs niet (Cito, 2012). In het belang van een goede vergelijkbaarheid en een zo eerlijk mogelijke beoordeling is het, bij een verandering van systematiek die tot een verbetering van de opbrengstenbeoordeling moet leiden, nodig om ook dit punt te tackelen zodat van alle scholen vergelijkbare gegevens beschikbaar zijn.

Overgangsfase

Het is de vraag of het de moeite waard is om een nieuwe correctiesystematiek te ontwikkelen die alle scholen administratief zwaarder belast, terwijl op voorhand bekend is dat ook deze systematiek niet optimaal corrigeert en bovendien zorgt voor een verlies aan transparantie. Voor een optimale correctie zijn namelijk leerwinstbepalingen op leerlingniveau nodig, die naast een verplichte eindtoets ook

een verplichte begin- of tussentoets (of meerdere tussentoetsen) vergen. Daarnaast zijn ook bij een leerwinstmodel betere gegevens nodig over de opleiding van ouders dan wat nu in het leerlinggewicht verwerkt is; aanpassing van de gegevensbevraging bij scholen is ook dan zeker gewenst. Het ligt echter volgens de inspectie het meest voor de hand om wijzigingen in de systematiek pas door te voeren als ook een leerwinstbepaling plaats kan vinden. Gekoppeld aan een meer gedetailleerde bevraging van scholen over de opleiding van ouders kan dan naar verwachting een grote stap voorwaarts gezet worden. Tot die tijd volstaat de huidige aanpak, zeker omdat scholen nooit alleen op basis van de onvoldoende opbrengsten zwak of zeer zwak kunnen worden.

In gesprekken over de resultaten van dit onderzoek is de mogelijkheid geopperd om voor scholen die als onvoldoende uit de huidige beoordelingssystematiek komen, aanvullend nog een ander model te hanteren met meer correctievariabelen, zoals model 4. Hiervoor opteert de inspectie niet, omdat resultaten voor scholen minder inzichtelijk worden. De uitkomst (wel of niet bijstelling van het oordeel) zal niet op voorhand duidelijk zijn voor scholen en daarom vaak onbevredigend.

Een beter alternatief lijkt in te zetten op een professionele dialoog van schoolbesturen met de inspectie. Als besturen van oordeel zijn dat op een school die als onvoldoende uit de risicoanalyse komt sprake is van bijzondere omstandigheden in de leerlingpopulatie, die afwijken van de beslisregels door de inspectie rechtvaardigen, kunnen zij hun visie beargumenteerd met de inspectie bespreken. De inspectie kan dan in haar oordeel eventueel beredeneerd afwijken van de beslisregels. Deze aanpak past in de systematiek waarin schoolbesturen verantwoordelijk zijn voor de kwaliteit van het onderwijs in hun scholen en zich daarover verantwoorden.

Literatuur

- Cfi (2008). *Nieuwe gewichtenregeling basisonderwijs*. Zoetermeer: Cfi.
- Cito (2011). *Terugblik en resultaten 2011. Eindtoets basisonderwijs groep 8*. Arnhem: Cito.
- Cito (2012). *Terugblik en resultaten 2012. Eindtoets basisonderwijs groep 8*. Arnhem: Cito.
- Dijkstra, A.B., Karsten, S., Veenstra, R., & Visscher, A.J. (2001). *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties*. Assen: Van Gorcum.
- Driessen, G., Mulder, L., & Roeleveld, J. (2012). *Cohortonderzoek COOL 5-18. Technisch rapport basisonderwijs, tweede meting 2010/11*. Nijmegen/Amsterdam: ITS/Kohnstamm Instituut.
- Godlieb, H. (2008). De weging gewogen. <http://www.deweginggewogen.nl>
- OCW (2011a). *Brief aan de Tweede Kamer betreffende de opbrengstenbeoordeling in het basisonderwijs, 12 december 2011*. Den Haag: OCW.
- OCW (2011b). *Brief aan de Tweede Kamer betreffende pilot leerwinst en toegevoegde waarde in het primair onderwijs*. Den Haag: OCW.
- OCW (2012a). *Brief aan de Tweede Kamer betreffende de verkenning differentiatie opleidingsniveau ouder(s) bij beoordeling opbrengsten primair onderwijs, 3 april 2012*. Den Haag: OCW.
- OCW (2012b). *Regeling van de Minister van Onderwijs, Cultuur en Wetenschap van 25 april 2012, nr WJZ/390987 (10201), houdende wijziging van de Regeling leerresultaten PO in verband met de invoering van een nieuwe systematiek voor de beoordeling van de Eindtoets Basisonderwijs van Cito met ingang van de leerresultaten die zijn behaald over schooljaar 2011-2012*. *Staatscourant*, 2012, nr. 9019, 9 mei 2012.
- Onderwijsraad (2003). *Wat scholen toevoegen*. Den Haag: Onderwijsraad.
- Roeleveld, J. (2003). *Herkomstkenmerken en begintoets, secundaire analyses op het PRIMA-cohortonderzoek*. Amsterdam: SCO-Kohnstamm Instituut.
- Roeleveld, J., Mooij, T., Fettelaar, D., Ledoux, G. (2011). *Correctiefactoren bij opbrengstmaten in het primair onderwijs*. Amsterdam/Nijmegen: Kohnstamm Instituut/ITS.

Bijlage 1 Geraadpleegde deskundigen

Dr. A. Beguin	Wetenschappelijk directeur Cito, Arnhem
Prof. dr. R.J. Bosker	Hoogleraar Onderwijskunde, Rijksuniversiteit Groningen
Prof. dr. W.J.C.M. van de Grift	Hoogleraar Onderwijskunde, Rijksuniversiteit Groningen
Prof. dr. G.W. Meijnen	Emeritus hoogleraar Onderwijskunde, Universiteit van Amsterdam
Dr. J. Roeleveld	Senior wetenschappelijk onderzoeker Kohnstamm Instituut, Amsterdam

Reactie externe deskundige en PO-raad op het rapport

De Inspectie van het Onderwijs heeft de data-analyses besproken met de heren Beguin, Roeleveld en Meijnen. Zij geven aan in te stemmen met de door de inspectie gehanteerde werkwijze/procedure en steunen de conclusie dat de verbetering bij het beoordelen van opbrengsten in het basisonderwijs door nadere differentiatie van het opleidingsniveau van ouders niet opweegt tegen de grotere administratieve lasten die daarmee bij het basisonderwijs worden neergelegd.

Het conceptrapport is voorgelegd aan de heren Bosker en Van de Grift. Hun opmerkingen zijn zoveel mogelijk in het rapport verwerkt.

Het conceptrapport is tevens met de PO-raad besproken; de raad heeft daarover de volgende verklaring afgegeven.

Het huidige systeem rondom het corrigeren van de eindopbrengsten is niet optimaal. Dit blijkt ook uit het (geringe) percentage verklaarde variantie waarnaar in het rapport wordt verwezen. De PO-Raad realiseert zich dat een wijziging in de systematiek leidt tot een betere schatting van de bijdrage van basisscholen aan prestaties. Echter dit verschil weegt, vanwege de blijvende imperfectie van het systeem, niet op tegen de extra lasten die dit voor het onderwijs met zich meebrengt. Deze extra lasten vindt de PO-Raad niet wenselijk, mede gezien de zeer kleine overhead op de scholen. De PO-Raad onderschrijft dan ook de insteek om meer ruimte te creëren voor een professionele dialoog tussen de schoolbesturen en de Inspectie als het gaat om de beoordeling van eindopbrengsten van scholen. De PO-Raad zal zich inspannen om de kwaliteit van deze professionele dialoog te versterken, door bij zijn leden hier aandacht voor te vragen. Ook de Inspectie heeft aangegeven hier binnen de organisatie aandacht aan te besteden. De PO-Raad hecht eraan dat er begin 2014 een evaluatieonderzoek plaatsvindt met de vraag of de schoolbesturen ook ervaren dat er meer ruimte is voor deze professionele dialoog.

Bijlage 2 **Ondergrenzen, gemiddelde scores en bovengrenzen, naar percentage gewogen leerlingen op school (huidige inspectiesystematiek voor het beoordelen van opbrengsten aan de hand van gemiddelde scores van basisscholen op de Eindtoets Basisonderwijs)**

% gewichten- leerlingen	Aantal scholen in populatie (1-10- 2010)	% scholen in populatie (1-10-2010)	Onder- grenzen	Gemiddelden	Boven- grenzen
0	601	9	535,2	537,2	539,2
1	357	5	535,1	537,1	539,1
2	357	5	534,9	536,9	538,9
3	395	6	534,8	536,8	538,8
4	372	5	534,6	536,6	538,6
5	324	5	534,5	536,5	538,5
6	322	5	534,3	536,3	538,3
7	341	5	534,2	536,2	538,2
8	321	5	534	536	538
9	298	4	533,9	535,9	537,9
10	246	4	533,7	535,7	537,7
11	212	3	533,5	535,5	537,5
12	196	3	533,4	535,4	537,4
13	207	3	533,2	535,2	537,2
14	172	3	533,1	535,1	537,1
15	147	2	532,9	534,9	536,9
16	146	2	532,8	534,8	536,8
17	119	2	532,6	534,6	536,6
18	130	2	532,4	534,4	536,4
19	112	2	532,3	534,3	536,3
20	83	1	532,1	534,1	536,1
21	95	1	532	534	536
22	86	1	531,8	533,8	535,8
23	65	1	531,6	533,6	535,6
24	73	1	531,5	533,5	535,5
25	52	1	531,3	533,3	535,3
26	63	1	531,2	533,2	535,2
27	55	1	531	533	535
28	42	1	530,9	532,9	534,9
29	47	1	530,7	532,7	534,7
30	31	0	530,6	532,6	534,6
31	32	0	530,4	532,4	534,4
32	32	0	530,3	532,3	534,3
33	38	1	530,1	532,1	534,1
34	21	0	530	532	534
35	31	0	529,9	531,9	533,9

36	28	0	529,7	531,7	533,7
37	21	0	529,6	531,6	533,6
38	25	0	529,4	531,4	533,4
39	26	0	529,3	531,3	533,3
40	19	0	529,2	531,2	533,2
41	23	0	529	531	533
42	22	0	528,9	530,9	532,9
43	16	0	528,8	530,8	532,8
44	20	0	528,7	530,7	532,7
45	13	0	528,5	530,5	532,5
46	19	0	528,4	530,4	532,4
47	24	0	528,3	530,3	532,3
48	26	0	528,2	530,2	532,2
49	19	0	528,1	530,1	532,1
50	14	0	528	530	532
51	19	0	527,9	529,9	531,9
52	16	0	527,8	529,8	531,8
53	18	0	527,7	529,7	531,7
54	30	0	527,6	529,6	531,6
55	19	0	527,5	529,5	531,5
56	15	0	527,4	529,4	531,4
57	9	0	527,4	529,4	531,4
58	17	0	527,3	529,3	531,3
59	14	0	527,2	529,2	531,2
60	9	0	527,2	529,2	531,2
61	17	0	527,1	529,1	531,1
62	16	0	527	529	531
63	15	0	527	529	531
64	11	0	526,9	528,9	530,9
65	9	0	526,9	528,9	530,9
66	7	0	526,8	528,8	530,8
67	8	0	526,8	528,8	530,8
68	7	0	526,8	528,8	530,8
69	5	0	526,8	528,8	530,8
70	8	0	526,7	528,7	530,7
71	3	0	526,7	528,7	530,7
72	3	0	526,7	528,7	530,7
73	8	0	526,7	528,7	530,7
74	6	0	526,7	528,7	530,7
75	2	0	526,7	528,7	530,7
76	2	0	526,7	528,7	530,7
77	2	0	526,7	528,7	530,7
78	2	0	526,7	528,7	530,7
79	3	0	526,7	528,7	530,7
80	2	0	526,7	528,7	530,7
81	1	0	526,7	528,7	530,7
82	3	0	526,7	528,7	530,7
83	0	0	526,7	528,7	530,7
84	0	0	526,7	528,7	530,7
85	4	0	526,7	528,7	530,7

86	0	0	526,7	528,7	530,7
87	0	0	526,7	528,7	530,7
88	1	0	526,7	528,7	530,7
89	1	0	526,7	528,7	530,7
90	0	0	526,7	528,7	530,7
totaal	6848	100			