

Kwaliteitsbeoordeling van scholen primair onderwijs: Het correctiemodel van de inspectie vergeleken met alternatieve modellen¹

T. Mooij, J. Roeleveld, D. Fettelaar en G. Ledoux

Samenvatting

Bij de beoordeling van de kwaliteit van scholen primair onderwijs gebruikt de inspectie met name de scores van leerlingen op de Cito-Eindtoets Basisonderwijs. Ter correctie van aanvangsverschillen van leerlingen wordt per school het geaggregeerd leerlinggewicht benut in de vorm van een globale score van de opleiding van de ouders. De onderzoeksvraag is gericht op identificatie van bruikbare andere correctiefactoren alsmede bepaling van de effecten van alternatieve correctiefactoren op de kwaliteitsbeoordeling. Ter beantwoording worden reeksen correctiemodellen ontworpen die zijn gebaseerd op enerzijds zeven leerlingkenmerken en anderzijds drie soorten multiniveau regressie-analyse. De onderzoeksmethode is secundaire analyse van gegevens verkregen met het landelijke COOL⁵⁻¹⁸ cohortonderzoek en data die worden beheerd door het Centraal Bureau voor de Statistiek. In de multiniveau analyses worden 402 scholen en 8.561 leerlingen betrokken. De resultaten demonstreren dat twee correctiefactoren (gedetailleerde opleiding ouders; etniciteit) relatief betere resultaten geven dan die welke worden verkregen via de huidige correctie met behulp van leerlinggewicht. De conclusie is dat, hoewel het onderzoek beperkingen kent, verbeteringen mogelijk zijn in de wijze waarop de inspectie de leeropbrengsten van scholen beoordeelt.

1 Inleiding

Een vrij algemene benadering ter inschatting van de kwaliteit van scholen is de bepaling van de 'toegevoegde waarde' of 'value-added' in de vorm van leeropbrengsten (Onderwijsraad, 2006). De Organisation for Economic Co-operation and Development (OECD) definieert toegevoegde waarde als 'the contribution of a school to students'

progress towards stated or prescribed education objectives (e.g. cognitive achievement). The contribution is net of other factors that contribute to students' educational progress' (OECD, 2008, p. 17). De OECD stelt tevens dat bij de bepaling van toegevoegde waarde van een school tenminste een begin- en eindmeting van leerlinggegevens nodig zijn. In dat geval kan worden vastgesteld hoeveel de leerlingen in leerprestaties vooruit gaan ('groeien') tijdens hun verblijf op school.

In de huidige systematiek ter beoordeling van de kwaliteit van Nederlandse scholen in het primair onderwijs gebruikt de Inspectie van het Onderwijs met name het gemiddelde eindresultaat van leerlingen op de school. Dit in de vorm van schoolscores op de Eindtoets Basisonderwijs van het Cito; deels zijn hierbij ook tussenresultaten op Cito-toetsen van belang (Inspectie van het Onderwijs, 2010). In het beoordelingsproces wordt aanvullend een zo adequaat mogelijke inschatting gemaakt van de schoolbijdrage aan de leerprestaties. Dit wil zeggen dat in het toezichtkader, conform de standaard werkwijze, rekening wordt gehouden met specifieke kwaliteitsaspecten en indicatoren ter beoordeling van de schoolresultaten (zie de betreffende specificaties op <http://cdn.ikregeer.nl/pdf/stert-2011-15145.pdf>). Tevens houdt de inspectie rekening met de (aanvangs)verschillen tussen de leerlingen via geaggregeerde schoolkenmerken. Dit in de vorm van het percentage leerlingen in de schoolpopulatie met gewicht 0.3 en 1.2 zoals die door het Cito bij de rapportages over de Eindtoets worden gebruikt. De inspectie beschikt niet over data waarmee een meer dynamische, op groei gerichte beoordelingssystematiek vanaf het begin van de schoolloopbaan mogelijk is.

De OECD noemt een benadering zoals de inspectie die volgt een 'contextualised attainment model' (OECD, 2008, p. 15). Hierin kan, zo mogelijk achteraf, worden gecorrigeerd voor kenmerken die de schoolloopbaan van

leerlingen kunnen beïnvloeden. Omdat er diverse mogelijkheden zijn te corrigeren voor de aanvangsverschillen tussen leerlingen, wilde de inspectie een onderzoek laten doen naar de waarde van de door haar gehanteerde correctiemethodiek en de (on)mogelijkheden van alternatieve correctiefactoren. De algemene onderzoeksvraag werd geformuleerd als: *Wat zijn bruikbare (andere) correctiefactoren bij de bepaling van de eindopbrengsten van basisscholen, wat zijn de effecten van alternatieve correctiefactoren voor de beoordeling van scholen, en wat zijn de effecten van correcties op verschillende niveaus?*

Ter beantwoording van de onderzoeksvraag werd besloten de huidige systematiek en andere mogelijkheden empirisch te verkennen. Dit door middel van secundaire analyses op de hiervoor meest geschikte data van de landelijke cohortgegevens uit het primair onderwijs (vgl. Driessen, Mulder, Ledoux, Roeleveld, & Van der Veen, 2009) en daarnaast aanvullende registerdata van het Centraal Bureau voor de Statistiek (CBS). In overleg met een begeleidingsgroep van de inspectie hebben wij verschillende reeksen modellen ontworpen waarin kenmerken van leerlingen op individueel niveau en op schoolniveau worden gebruikt om de 'ruwe' uitkomsten op de Eindtoets Basisonderwijs te corrigeren (zie Roeleveld, Mooij, Fetteelaar, & Ledoux, 2011). Een eerste beoordelingsmodel is de correctie die de inspectie tot op heden hanteert. Analyseresultaten uit dit model worden vervolgens vergeleken met resultaten van analyses van series andere correctiemodellen. Ter eenduidige vergelijking van de resultaten worden de analyses zoveel mogelijk uitgevoerd op gegevens van eenzelfde groep scholen.

2 Beoordelingssystematiek: theoretisch kader

2.1 Modellering van correctiefactoren

Er zijn diverse soorten invloeden mogelijk bij de realisatie van leerlingprestaties op scholen. Te noemen zijn met name kenmerken van de persoon(lijkheid) van de leerling, het gezin waaruit deze komt, de school zelf en de buitenschoolse situatie van de leerlingen

respectievelijk de wijk waarin de school staat (vgl. Driessen, Mooij, & Doesborgh, 2007; Onderwijsraad, 2006; Roeleveld, Driessen, Ledoux, Cuppen, & Meijer, 2011). In de context van de huidige vraagstelling inclusief de beschikbare gegevens ter beantwoording hiervan gaat het er vooral om dat de beoordeling van scholen kan variëren op grond van de relevantie van deze diverse soorten invloeden. Het doel daarvan is de beoordeling zo eerlijk mogelijk, zo zuinig mogelijk én, voor scholen zelf, zo begrijpelijk mogelijk te doen zijn.

In de huidige onderzoekssituatie maken wij gebruik van schoolgegevens die aan het eind van het primair onderwijs zijn verzameld in het kader van het landelijke COOL⁵⁻¹⁸ cohortonderzoek (Driessen et al., 2009). Leerprestaties kunnen dan op verschillende wijzen worden gecorrigeerd voor kenmerken van de leerlingen in deze scholen. Rekening houdend met de beschikbare gegevens, kan de vergelijking van mogelijke correcties in de beoordelingssystematiek van scholen worden uitgewerkt via twee soorten operationaliserings- ofwel modellen. De eerste soort is gericht op correctie via individuele kenmerken van leerlingen (leerlingkenmerken of -covariaten) en de tweede op correctie met behulp van gemiddelde of compositiekenmerken op schoolniveau. Deze twee typen kenmerken kunnen ook worden gecombineerd, waarbij verschillende (combinaties van) correctiefactoren kunnen worden gebruikt. Daarnaast kunnen modellen worden opgesteld volgens een klassieke 'platte' of via multiniveau regressie-analyse. In tegenstelling tot in multiniveau analyse, wordt in platte modellen geen rekening gehouden met het genest zijn van leerlingen binnen scholen. Deze platte variant hebben wij op verzoek van de begeleidingscommissie meegenomen in het oorspronkelijke onderzoek (in dit artikel stellen wij deze modellering slechts summier aan de orde). De diverse mogelijkheden resulteren in een matrix met correctiemodellen die schematisch wordt weergegeven in Tabel 1.

De mogelijke correctiefactoren zijn covariaten die achtergrondkenmerken betreffen van leerlingen: zie de linker kolom van Tabel 1. Het gaat respectievelijk om leerlinggewicht (A), opleiding ouders (B), etniciteit

Tabel 1

Alternatieve modellen ter beoordeling van schoolkwaliteit, geordend naar covariaat en type

Niveau operationalisatie:	Schoolcompositie-kenmerken		Individuele kenmerken		School- + individ. kenmerken	
Niveau modellering:	Multi-niveau*	Plat	Multi-niveau*	Plat	Multiniveau*	
Covariaat:	Model:	1	2	3	4	5
A. Leerlinggewicht	A1	A2	A3	A4	A5	
B. Opleiding ouders	B1	B2	B3	B4	B5	
C. Etniciteit	C1	C2	C3	C4	C5	
ABC combinatie					ABC5	
D. Huishoudinkomen	D1	D2	D3	D4	D5	
E. % lage inkomens	E1	E2	E3	E4	E5	
F. % niet-westers allochtoon	F1	F2	F3	F4	F5	
G. % uitkeringen	G1	G2	G3	G4	G5	
ABCDEFG combinatie					ABCDEFG5	

* Twee niveau's: leerling- en schoolniveau; random intercept model.

(C), huishoudinkomen van het gezin van de leerling (D), en drie kenmerken van de wijk waarin een leerling woont: % lage inkomens (E), % niet-westerse allochtonen (F), en % uitkeringsgerechtigden (G). Per rij worden de correctiespecificaties bij elke leerlingcovariaat weergegeven. De zeven covariaten zijn per school geaggregeerd: zie de modellen 1 en 2 met schoolcompositiekenmerken in Tabel 1. De covariaten zijn ook individueel, per leerling, bepaald en gemodelleerd: zie de modellen 3 en 4. De schoolcompositiekenmerken kunnen worden opgenomen in een klassieke platte regressievergelijking zonder onderscheid in niveau's: zie modellen A2 tot en met G2. Rekening houden met de hiërarchische datastructuur doen we volgens een random intercept model waarbij intercepten variëren tussen scholen: zie modellen A1 tot en met G1. Hetzelfde geldt voor de individuele leerlingkenmerken. Deze kunnen als onafhankelijke variabele gemodelleerd worden met een klassieke regressie (modellen A4 tot en met G4) en volgens een bepaling met twee niveau's (modellen A3 tot en met G3). Tevens onderscheiden we nog een vijfde soort modellen die de combinatie representeert van individuele én schoolcompositiekenmerken volgens een multiniveau modellering (A5 tot en met G5).

Elke cel in Tabel 1 verwijst naar een specifiek submodel dat wordt gekarakteriseerd door middel van een covariaat (letter) ge-

volgd door een cijfer. Ook zijn diverse combinatiemodellen mogelijk, zoals bijvoorbeeld ABC5 of ABCDEFG5. Model A1 is het beoordelingsmodel dat de inspectie momenteel hanteert. Dit model is gebaseerd op de door het Cito gerapporteerde, voor gemiddeld leerlinggewicht gecorrigeerde schoolscore op de Eindtoets Basisonderwijs. In het inspectiemodel wordt daarmee gecorrigeerd voor op schoolniveau geaggregeerde (gepercentreerde) leerlinggewichten. Het meest volledige model, met de meeste leerlingcovariaten conform een multiniveauanalyse, staat rechtsonder in Tabel 1. In dit model ABCDEFG5 zijn alle covariaten opgenomen, zowel op individueel niveau als geaggregeerd op schoolniveau. Bij dit laatste correctiemodel is, theoretisch, de kans het kleinst dat resterende schoolverschillen te maken hebben met verschillen in de leerlingbevolking van de scholen.

Tabel 1 maakt duidelijk dat een grote hoeveelheid modellen en daaruit resulterende classificaties van scholen onderling kunnen worden vergeleken. Gezien de onderzoeksvraagstelling is het vooral relevant na te gaan of er, in vergelijking met model A1, aantoonbaar betere modellen zijn die tot een vergelijkbare classificatie van scholen leiden als het meest volledige correctiemodel ABCDEFG5, maar die tegelijk minder complex zijn dan dit laatste model. In overleg met de begeleidingsgroep van de inspectie werd be-

sloten het onderzoek te richten op de modellen 1, 2 en 5 (vgl. Roeleveld et al., 2011). De platte regressiemodellen 2 werden in dit onderzoek meegenomen omdat een dergelijk model voor scholen mogelijk meer inzichtelijk is dan de (statistisch geavanceerdere) multiniveau modellen 1 en 5. Omdat bekend is dat deze laatste modellen nauwkeuriger schattingen van schooleffecten opleveren, concentreren wij ons in dit artikel echter op de modellen 1 en 5.

2.2 Criteria ter classificatie van scholen

Bij elk onderzoeksmodel kan per school het residu (de afwijkingsscore op basis van regressie-analyse) worden bepaald. De waarde van het residu geeft aan hoeveel de op die school behaalde gemiddelde score op de Cito-Eindtoets afwijkt van de score die door het betreffende correctiemodel statistisch wordt 'voorspeld', ofwel verwacht op basis van kenmerken van leerlingen van deze school. De waarden van de schoolresiduen geven, voor elk model, gelegenheid tot ordening van scholen in termen van grotere of kleinere, positieve of negatieve afwijkingen van de verwachte score van die school. Vanuit het perspectief van de inspectie gaat het daarbij in het bijzonder om het opsporen van scholen die onder de norm blijven: scholen met grote negatieve residuen. Dit duidt op lagere gemiddelde opbrengsten dan op basis van het betreffende correctiemodel kan worden verwacht.

Een eerste criterium ter beoordeling van de verschillende correctiemodellen bestaat dan in de berekening van correlaties tussen de residuen van die modellen. De hoogte van die correlaties geeft een globaal beeld van veranderingen in de ordening van scholen, afhankelijk van de correctiemodellen. Een tweede criterium is de bepaling van hoeveel scholen bij het ene correctiemodel als 'onder de norm' worden getypeerd en bij het andere model niet. Hierbij is de inspectie vooral geïnteresseerd in de wisseling tussen wel/niet onder de norm omdat alleen dat consequenties kan hebben voor het totale schooloordeel. Conform de standaardwerkwijze van de inspectie is besloten een halve standaarddeviatie als grens voor de typering van de leerprestaties op een school als 'onder de

norm' te gebruiken. Dit houdt in dat het schoolresidu tenminste een halve standaarddeviatie (SD) lager is dan het algemeen gemiddelde van de residuen (dat bij deze residuen 0 is). Bij elk correctiemodel wordt daarom de standaarddeviatie van de schoolresiduen bepaald en wordt een indeling van scholen gemaakt in:

- onder de norm ($1/2$ SD of meer onder het gemiddelde);
- normaal, op de norm (tussen + of $- 1/2$ SD van het gemiddelde);
- boven de norm ($1/2$ SD of meer boven het gemiddelde).

Verskillende correctiemodellen van leeropbrengsten leiden tot verschillende beoordelingsclassificaties van scholen in termen van 'onder de norm', 'normaal' en 'boven de norm'. In de context van schoolbeoordeling kunnen deze classificaties met elkaar worden vergeleken. Zo kan helder worden in hoeverre eventuele verschuivingen in de positie van scholen afhankelijk zijn van de gehanteerde correctiemethode. Een derde criterium is dan wat de kenmerken zijn van scholen waarvan de opbrengsten in één of meer modellen als 'onder de norm' worden beoordeeld.

3 Methode

3.1 Secundaire analyse van COOL- en CBS-gegevens

In de secundaire analyses worden ten eerste gegevens gebruikt van leerlingen uit scholen die participeerden aan de eerste meting van het COOL⁵⁻¹⁸ cohortonderzoek (verder COOL genoemd; zie Driessen et al., 2009). In dit longitudinaal cohortonderzoek worden, om de drie jaar, gegevens verzameld bij een groot aantal scholen en leerlingen. In de eerste meting (schooljaar 2007/2008) participeerden leerlingen uit de groepen 2, 5 en 8 van het primair onderwijs en leerjaar 3 van het voortgezet onderwijs. Aan deze meting hebben 550 basisscholen meegedaan, met ruim 11.000 leerlingen in groep 8. De steekproef scholen bestaat uit 400 scholen die samen een representatief beeld van het Nederlandse primair onderwijs geven, alsmede een aanvullende steekproef van 150 scholen met veel achterstandsleerlingen. In onze ana-

lyses gebruiken wij de gegevens van al deze scholen. Van uiteindelijk 402 van deze COOL-scholen zijn scores op de Cito-Eindtoets Basisonderwijs beschikbaar en zijn tevens voldoende gegevens betreffende de leerlingcovariaten aanwezig.

Een probleem bij de Eindtoets is dat een deel van de leerlingen niet wordt getoetst. Dat geldt bijvoorbeeld voor leerlingen die uitstromen naar Leerwegerondersteunend onderwijs (LWOO) of Praktijkonderwijs (PRO). De inspectie kent aan leerlingen bij wie de score op de Eindtoets ontbreekt én naar het LWOO gaan, de score 517 toe; niet-deelnemers aan de Eindtoets die naar het PRO gaan, krijgen geen score toegekend (Inspectie van het Onderwijs, 2010). De score 517 is ongeveer het gebruikelijke niveau van een leerling met advies VMBO-basis (Vorbereidend Middelbaar Beroepsonderwijs-basis). Ook in het COOL-onderzoek is van 5.1% van de leerlingen geen score op de Cito-Eindtoets bekend. Wel zijn veelal scores op andere Cito-toetsen uit groep 8 verkregen. Op basis van deze andere toetsscores en enkele achtergrondkenmerken zijn, via de multiële imputatie-procedure van SPSS, schattingen verkregen van scores op de Eindtoets voor de leerlingen zonder score. Leerlingen waarvan geen Cito-Eindtoetsscore bekend is, scoren gemiddeld 524.6 op de Eindtoets; leerlingen waarvan deze score wel bekend is, scoren gemiddeld 532.9 (totale N leerlingen: 8.561). Circa de helft van de leerlingen zonder score op de Eindtoets heeft helemaal geen uitstroominformatie, terwijl de overigen ook wel havo- of vwo-advies krijgen. Dit duidt op omissies bij het verstrekken van de informatie door de school. Via de statistische imputatie van de scores op de Eindtoets worden alle leerlingen betrokken in de analyses van mogelijke correctiemodellen.

Bovendien worden gegevens benut uit drie databestanden die worden beheerd door het Centraal Bureau voor de Statistiek (CBS, 2009, 2010a, 2010b). De bestanden bevatten informatie ter operationalisatie van de overige vier leerlingkenmerken of -covariaten (zie Tabel 1). Het eerste CBS-bestand betreft data van de gemeentelijke basisadministratie (GBA). Dit zijn gegevens van de gehele Nederlandse populatie personen en huishoudens

die staat geregistreerd bij gemeenten. Wij hebben een selectie benut van persoons- en huishoudkenmerken met peildatum 25 januari 2008; het huishoudnummer is bepaald op 1 januari 2008. Een volgend CBS-bestand is het integraal inkomensbestand (IIB), eveneens uit 2008. Dit bevat ook alle personen in Nederland. De inkomensgegevens zijn met name afkomstig van de belastingdienst en daarnaast van andere administratieve instanties. Tenslotte is het sociaal statistisch bestand (SSB) gebruikt. Dit betreft een verzamelingbestand met alle personen in Nederland waarvan gegevens zijn verzameld uit een veelvoud aan registerbestanden en enquêtes. Ook hier is uitgegaan van de stand van zaken op 25 januari 2008.

3.2 Leerlingcovariaten en schoolgemiddelden

Leerlingkenmerken COOL

De leerlingkenmerken A, B en C uit Tabel 1 worden rechtstreeks uit het COOL-bestand van groep 8 afgeleid. Het gaat om: A: *leerlinggewicht*; B: *opleiding van de ouders*; en C: *etnische herkomst*. Leerlinggewicht en opleiding ouders zijn gemeten volgens de criteria in de nieuwe gewichtenregeling (Mulder, Roeleveld, Van der Veen, & Vierke, 2005; Mulder & Vierke, 2007; Roeleveld, Mooij et al., 2011). In deze gewichtenregeling wordt pas een gewicht toegekend als beide ouders ten hoogste het niveau van lager beroepsonderwijs (lbo) hebben bereikt. De resultaten zijn als volgt.

A. *Leerlinggewicht*: van de leerlingen (n = 8.561) in groep 8 krijgt 66.1% geen gewicht; 17.5% heeft gewicht 0.3; 15.2% verkrijgt gewicht 1.2; en 1.2% is onbekend.

B. *Opleiding ouders*: is ingedeeld conform de nieuwe gewichtenregeling. Bij ouders die beiden meer dan lager beroepsonderwijs hebben gehad, is onderscheid gemaakt tussen wel of geen hoger beroepsonderwijs (hbo) of wetenschappelijk onderwijs (wo) hebben afgerond. Het resultaat staat in Tabel 2.

C. *Etnische herkomst*: hier is gekeken naar de geboortelanden van beide ouders. Het resultaat is de indeling zoals weergegeven in Tabel 3.

Tabel 2

Opleiding van de ouders van leerlingen uit groep 8 (percentages)

Beide ouders max. lbo; tenminste 1 max. lager onderwijs	15.2%
Beide ouders max. lbo	17.5%
Één ouder max. lbo, de ander meer	22.1%
Beide ouders meer dan lbo; geen hbo of wo	21.3%
Beide ouders meer dan lbo; minstens één heeft hbo of wo	21.2%
Onbekend	2.7%
Aantal leerlingen	8.561

Tabel 3

Etnische herkomst van leerlingen uit groep 8 (percentages)

Nederlands	67.6%
Overig westers	2.9%
Turks	7.9%
Marokkaans	6.7%
Surinaams/Antilliaans	3.9%
Overig niet-westers	6.1%
Gemengd Nederlands/allochtoon	4.5%
Onbekend	0.5%
Aantal leerlingen	8.561

Leerlingenmerken CBS

Op leerlingniveau zijn de COOL-gegevens gekoppeld aan diverse CBS-gegevens die werden geïdentificeerd via naam-adres-woonplaatsinformatie (NAW-gegevens ofwel RIN-nummer). Het betreft ten eerste het volgende leerlingenmerk (vgl. Tabel 1):

D: het gestandaardiseerd huishoudinkomen per leerling. Tot welk huishouden een leerling behoort, is vastgesteld via de GBA. Met behulp van het IIB is dan het gestandaardiseerde huishoudinkomen te bepalen. Dit is het 'besteedsbaar huishoudinkomen' ofwel het bruto huishoudinkomen (primair inkomen, inkomen uit uitkeringen, verzekeringen en sociale voorzieningen, gebonden overdrachten (bijvoorbeeld huurtoeslag) en inkomensoverdrachten (bijvoorbeeld alimentatie)) minus betaalde inkomensoverdrachten, premies inkomensverzekeringen en belasting op inkomen en vermogen. Het resulterende inkomen is gestandaardiseerd, dit wil zeggen gecorrigeerd voor grootte en samenstelling van het huishouden. Het huishoudinkomen indiceert dan de daadwerkelijke (gedeelde) welvaart van het huishouden waartoe een leerling behoort. Aan 94% van de leerlingen met geldige waarden uit het COOL-bestand kon een RIN nummer worden toegekend. Via deze weg was het mogelijk de postcode-gebaseerde gegevens toe te

kennen aan 93,9% van de leerlingen. Aan in totaal 86,8% van de leerlingen kon het gestandaardiseerde huishoudinkomen worden toegevoegd.

Het CBS heeft de viercijferige postcode van het woonadres van personen toegevoegd aan het GBA-bestand. Via de postcode hebben wij GBA-gegevens op postcodeniveau weer gekoppeld aan de leerlingen. Het gaat om drie contextuele leerlingenmerken:

E: Percentage lage inkomens per postcode. Het CBS heeft voor 2008 de 'lage inkomensgrens' vastgesteld op euro 11.020 per jaar, of euro 920 per maand, voor een één-persoonshuishouden (Centraal Bureau voor de Statistiek, 2009). Omdat we de beschikking hebben over kenmerk D, het gestandaardiseerde huishoudinkomen, is deze informatie voldoende om te bepalen of het huishoudinkomen van een leerling al dan niet een 'laag' inkomen is. Vervolgens hebben we deze bepaling per huishouden geaggregeerd naar het postcodeniveau zodat we de beschikking hebben over het percentage lage inkomens per postcodegebied. Huishoudens waarvan het hoofd scholier of student is, zijn hierbij niet meegeteld (dit is bepaald via de Sociaal Economische Categorie (SEC)).

F: Percentage niet-westerse allochtonen per postcode. Alle personen in de GBA zijn ingedeeld in twee categorieën: (a) autochtoon of westerse allochtoon, of (b) niet-westerse allochtoon. Bepalend hierbij is de CBS-definitie van herkomst: allochtonen zijn personen die zelf in het buitenland geboren zijn, of waarvan één, of allebei de ouders, in het buitenland geboren zijn; in de andere gevallen is men autochtoon. Niet-westerse allochtonen zijn zelf, of één van hun ouders is, afkomstig uit Europa (behalve Nederland en Turkije), Noord-Amerika, Indonesië, Japan, of Oceanië. Om het percentage niet-westerse

allochtonen per postcodegebied te bepalen, zijn deze gegevens geaggregeerd naar het postcodeniveau.

G: Percentage uitkeringsgerechtigden per postcode. In het SSB is de voornaamste inkomensbron van personen opgenomen via de SEC. Het percentage uitkeringsgerechtigden (uitkering als voornaamste inkomensbron) hebben we aan de hand hiervan bepaald. Het gaat om een uitkering in verband met arbeidsongeschiktheid, werkloosheid, bijstand of een ander soort uitkering (geen pensioen-uitkering, studiefinanciering etc.). Per postcode is het percentage inwoners bepaald dat afhankelijk is van een uitkering. Niet meegenomen in de berekening zijn kinderen, scholieren en studenten.

Tabel 4 bevat informatie afkomstig van het CBS. Gegeven worden de minimum- en maximumwaarde, het gemiddelde en de standaarddeviatie van deze vier leerlingcovariaten. Betrokken zijn de leerlingen die een geldige waarde hebben op alle variabelen ($n = 7.435$). Postcodekenmerken van de leerlin-

gen die hier vanwege ontbrekende waarden buiten vallen, zijn wel gebruikt voor het verkrijgen van de betreffende schoolkenmerken (via aggregatie van de leerlingkenmerken).

Schoolgemiddelden van de leerlingkenmerken

Ten behoeve van de diverse analyses zijn de leerlingkenmerken A tot en met G geaggregeerd naar schoolniveau. De univariate resultaten op schoolniveau vormen de schoolcompositiekenmerken; deze zijn weergegeven in Tabel 5.

3.3 Berekening van percentages variantie

Bij de multilevel (ML) modellen hanteren we een random intercept model (met schattingen van variantie op leerling- en schoolniveau). Hierbij wordt nagegaan hoeveel beide soorten varianties gereduceerd worden door invoer van verklarende variabelen op leerlingniveau. Een ML-model met een sterkere reductie dan in een ander ML-model is te

Tabel 4

Leerlingkenmerken gebaseerd op de CBS-gegevens (n leerlingen = 7.435)

	Min.	Max.	M	SD
Gestand. besteedbaar inkomen huish.	-90.193.0 ¹	791.000.0	21.200.1	17.365.1
% lage inkomens postcode	0.0	63.0	5.2	2.5
% niet-westerse allochtonen postcode	0.0	87.1	14.2	17.0
% uitkering postcode	1.6	35.7	12.2	5.6

¹ De berekening van het gestandaardiseerde huishoudinkomen door het CBS houdt in dat dit netto-inkomen negatief kan zijn, bijvoorbeeld bij zelfstandigen met verlies uit een onderneming.

Tabel 5

Schoolkenmerken gebaseerd op de naar school geaggregeerde leerlingkenmerken

	Min.	Max.	M	SD
<i>COOL-variabelen¹</i>				
A. % gewicht 0,3	0.0	80.0	18.7	15.4
% gewicht 1,2	0.0	100.0	16.0	20.1
B. Opleiding ouders	1.0	5.0	3.1	0.8
C. % allochtoon	0.0	100.0	29.0	32.6
<i>CBS-variabelen²</i>				
D. huishoudinkomen	-4.668.0 ³	72.890.9	20.395.0	6179.8
E. % lage inkomens (postcode)	1.6	12.8	5.3	2.13
F. % niet-westerse allochtonen (postcode)	0.0	72.0	13.5	15.4
G. % uitkeringen (postcode)	3.4	28.4	12.1	5.2

¹ COOL-bestand: n scholen = 402.

² CBS-bestanden: n scholen = 401.

³ De berekening van het gestandaardiseerde huishoudinkomen door het CBS houdt in dat dit netto-inkomen negatief kan zijn, bijvoorbeeld bij zelfstandigen met verlies uit een onderneming.

Tabel 6

Correlaties tussen model-residuen van leerlingcovariaten A t/m G, modellen 1 en 5

Covariaten COOL	Model	A: Leerlinggewicht		B: Opleiding ouders		C: Etniciteit		
		A1	A5	B1	B5	C1	C5	
A: Leerlinggewicht	A1							
	A5	.99						
B: Opleiding ouders	B1	.97						
	B5		.96	.99				
C: Etniciteit	C1	.88		.85				
	C5		.89		.85	.98		
D: Huishoudinkomen	Model	D: Huish.-inkomen		E: % lage inkomens		F: % n.w.-allochtoon		G: % uitkering
		D1	D5	E1	E5	F1	F5	G1
D: Huishoudinkomen	D1							
	D5	1.00						
E: % lage inkomens	E1	.86						
	E5		.86	1.00				
F: % n.w.-allochtoon	F1	.85		.97				
	F5		.85		.97	1.00		
G: % uitkeringen	G1	.86		.96		.95		
	G5		.86		.96		.95	1.00

prefereren omdat het risico dat de resterende schoolverschillen toe te schrijven zijn aan verschillen in leerlingkenmerken dan kleiner is.

4 Resultaten

4.1 Correlaties tussen modelresiduen

Een eerste criterium ter beoordeling van de verschillende correctiemodellen bestaat uit de correlaties tussen de residuen van modellen. De hoogte van de correlaties indiceert de veranderingen in de ordening van scholen, afhankelijk van de gebruikte correctiemodellen. Allereerst zijn modellen geanalyseerd waarin de COOL-leerlingcovariaten A, B en C uit Tabel 1 zijn opgenomen. Wij beperken ons, als gezegd, tot de modellen 1 en 5 van Tabel 1. Hierbij kijken we eerst naar de correlaties tussen modellen waarbij dezelfde covariaat verschillend gemodelleerd wordt, zoals: A1 met A5; tot en met G1 met G5. De correlatiecoëfficiënten tussen de modelresiduen staan in Tabel 6. De verschillende modellering van eenzelfde covariaat A tot en met G levert schoolresiduen op waarvan de correlaties variëren van 0.98 tot en met 1.00. Er is hier dus nauwelijks verschil.

Tabel 6 bevat ook de resultaten die corres-

ponderen met de analyse volgens eenzelfde soort model, bijvoorbeeld A1, B1 enzovoorts, maar met steeds andere covariaten. In dit opzicht doen zich relatief meer verschillen voor: de laagste correlatie is 0.85.

Aanvullend zijn nog correlaties berekend tussen het huidige inspectiemodel (A1) en enkele meer volledige modellen 5 (multilevel met covariaten op zowel individueel als op schoolniveau). De resultaten daarvan staan in Tabel 7 en zijn van dezelfde orde als die in Tabel 6.

4.2 Verklaarde varianties

Enkelvoudige modellen

Het percentage verklaarde variantie per model is een indicatie voor de afhankelijk-

Tabel 7

Correlaties tussen model-residuen A1 met meer volledige modellen 5

Correlaties tussen schoolresiduen	
A1 met ABC5	.93
A1 met EFG5	.88
A1 met DEFG5	.88
A1 met ABCD5	.92
A1 met ABCDEFG5*	.91

* Convergentie na 100 iteraties.

Tabel 8

Verklaarde varianties van multiniveau modellen met enkelvoudige regressie-analyse

Kenmerken multiniveau model:		Schoolcom-positie model 1	Individueel model 3	School- en indiv. model 5
Covariaten:	Niveau:			
A: Leerlinggewicht	school	44.9%	33.9%	44.6%
	leerling	0.0%	5.6%	5.7%
	totaal	6.6%	9.8%	11.4%
B: Opleiding ouders	school	49.0%	43.5%	48.9%
	leerling	0.0%	9.9%	9.9%
	totaal	7.2%	14.8%	15.6%
C: Etniciteit	school	34.7%	24.3%	34.3%
	leerling	0.0%	1.1%	1.2%
	totaal	5.1%	4.5%	6.0%
D: Huishoudinkomen	school	37.0%	10.2%	36.7%
	leerling	-0.1%	0.7%	0.7%
	totaal	5.4%	2.1%	6.1%
E: % lage inkomens	school	13.9%	11.9%	13.9%
	leerling	0.0%	0.0%	0.1%
	totaal	2.1%	1.8%	2.1%
F: % n.w.-allochtoon	school	19.7%	19.2%	19.6%
	leerling	0.0%	0.15%	0.2%
	totaal	3.0%	3.0%	3.1%
G: % uitkeringen	school	23.9%	23.0%	23.7%
	leerling	0.1%	0.2%	0.3%
	totaal	3.6%	3.6%	3.8%

heid van de eindresultaten van leerlingen op scholen ten opzichte van kenmerken van de leerlingbevolking op de scholen. Een hoger percentage betekent een beter voorspellend correctiemodel. Deze percentages kunnen verschillen afhankelijk van (1) de leerling- en schoolcovariaten die in de modellen zijn opgenomen en (2) de toepassing van multiniveau-modellering van het leerling- en schoolniveau: vgl. Tabel 1. Een eerste overzicht van percentages verklaarde variantie bevat enkelvoudige regressiemodellen ofwel de resultaten per leerlingcovariaat voor de modellen 1, 3 en 5: zie Tabel 8.

De resultaten in Tabel 8 laten zien dat de modellen 5, met de covariaat op zowel individueel als geaggregeerd niveau, vrijwel steeds meer variantie verklaren dan de modellen 1 of 3. De relatief meeste totaalvariantie (15.6%) wordt verklaard door model B5 (opleiding ouders op individueel- en schoolniveau). In het inspectiemodel A1 is leerlinggewicht wel opgenomen, maar zonder het onderscheid tussen middelbaar en hoger opgeleide ouders. Het bij A1 behorende percentage verklaarde variantie is 6.6%. In het geval dat deze leerlinggewichten ook op indivi-

duel niveau worden meegenomen zoals in model A5, neemt de verklaarde totale variantie toe tot 11.4%. Covariaat C (etnische herkomst) verklaart duidelijk minder (6.0%) dan opleiding ouders (B) of gewicht (A). Hetzelfde geldt voor covariaat D (huishoudinkomen) dat in totaal 6.1% verklaart. De drie variabelen afgeleid uit het postcodegebied waarin de leerling woont (E, F en G) verklaren alle minder variantie (2.1 – 3.8%) dan de modellen A – D.

Meervoudige modellen

In Tabel 9 zijn de verklaarde varianties opgenomen van modellen 1, 3 en 5 met diverse combinaties van de leerlingcovariaten. De gearceerde cellen geven aan dat er convergentieproblemen optreden (het model kan niet of moeilijk worden geschat, hoogstwaarschijnlijk vanwege de sterke samenhang tussen de variabelen). Ook het meest uitgebreide model ABCDEFG5 convergeert niet; de weergegeven resultaten van dit model zijn die van de schatting na 70.000 iteraties.

In Tabel 9 blijkt dat een meervoudig regressiemodel meer variantie kan verklaren, of beter kan voorspellen, dan het beste enkel-

Tabel 9

Verklaarde varianties van modellen met meervoudige regressie-analyse

Kenmerken multiniveau model:		Schoolcompositie model 1	Individueel model 3	School- en indiv. model 5
Covariaten:	Niveau			
ABCDEFG	school	53.7%	50.0%	55.1%
	leerling	0.0%	11.8%	11.7%
	totaal	****	17.5%	****
BC	school	52.6%	47.4%	52.6%
	leerling	0.0%	10.4%	10.4%
	totaal	7.7%	15.8%	16.6%
ABC	school	55.5%	47.4%	55.8%
	leerling	0.0%	10.4%	10.4%
	totaal	8.1%	15.8%	17.1%
ABCD	school	55.5%	48.7%	55.1%
	leerling	0.0%	11.7%	11.7%
	totaal	8.3%	17.2%	****
EFG	school	26.5%	25.1%	26.3%
	leerling	0.0%	0.2%	0.3%
	totaal	4.0%	3.9%	4.2%
DEFG	school	43.2%	30.6%	42.8%
	leerling	-0.1%	0.9%	0.9%
	totaal	6.4%	5.4%	7.2%
BCEFG	school	50.4%	47.8%	49.1%
	leerling	0.0%	11.4%	11.4%
	totaal	7.6%	16.9%	17.4%
BCDEFG	school	53.2%	48.5%	52.9%
	leerling	-0.1%	11.5%	11.5%
	totaal	7.9%	17.1%	****
AEFG	school	43.9%	43.5%	46.0%
	leerling	0.0%	6.9%	6.9%
	totaal	6.6%	12.4%	12.7%
ADEFG	school	49.8%	45.8%	51.1%
	leerling	0.0%	7.2%	7.2%
	totaal	7.4%	13.0%	13.8%

****: analyseresultaat wordt gekenmerkt door convergentieproblemen.

voudige model in Tabel 8. Een relatief goed model in Tabel 9, zonder convergentieproblemen, is model ABC5 met de covariaten leerlinggewicht, opleiding ouders en etnische herkomst. Dit model verklaart 17.1% van de totale variantie. Een ander relatief goed model is model BC (covariaten opleiding ouders en etnische herkomst). Dit model verklaart 16.6% van de variantie.

4.3 Verschuivingen in beoordeling van scholen

Op basis van de verschillende modellen kunnen de opbrengsten van scholen steeds ingedeeld worden als 'onder de norm', 'normaal', of 'boven de norm' (zie par. 2.2). Tabel 10 bevat in de bovenhelft de percentages scholen die, per leerlingcovariaat, tussen model-

len 1 en 5 een stabiele kwalificatie (onder de norm, normaal, boven de norm) hebben en de percentages scholen die wisselen van beoordelingskwalificatie (afhankelijk van covariaat en type modellering). Wat betreft de covariaten leerlinggewicht, opleiding ouders en etniciteit wisselt circa 5% van de scholen tussen de kwalificaties 'onder de norm' en 'normaal'; bij de overige covariaten is dit percentage (vrijwel) 0. Bij deze respectievelijke groepen covariaten wisselen vergelijkbare percentages tussen 'normaal' en 'boven de norm'.

De benedenhelft van Tabel 10 geeft de percentages scholen weer die, in eenzelfde model en met wisselende leerlingcovariaten, een stabiele ofwel wisselende kwalificatie krijgen. Afhankelijk van voor welke covaria-

Tabel 10

Percentages verschuivingen in schoolbeoordeling bij enkelvoudige modellen 1 en 5

	Stabiel onder norm	Wisseling onder norm / normaal	Stabiel normaal	Wisseling normaal / boven norm	Stabiel boven norm
<i>Per leerlingcovariaat</i>					
A. Leerlinggewicht	27,4	4,7	36,8	3,5	27,6
B. Opleiding ouders	26,9	5,2	35,3	5,0	27,6
C. Etniciteit	28,1	4,7	32,1	8,5	26,6
D. Huishoudinkomen	27,4	0,2	38,9	0,2	33,2
E. % lage inkomens modellen	26,9	0,0	42,4	0,0	30,7
F. % niet-west. all. modellen	28,2	0,0	41,1	0,0	30,7
G. % uitkeringen modellen	28,2	0,0	40,9	0,2	30,7
<i>Per model</i>					
Modellen 1 (A,B,C,ABC)	20,9	16,4	23,1	17,2	22,4
Modellen 5 (A,B,C,ABC)	20,6	17,4	20,6	21,1	20,1
Modellen 1 (D, E, F, G)	20,4	13,2	23,4	21,4	21,4
Modellen 5 (D, E, F, G)	20,4	13,0	23,9	21,2	21,4

ten wordt gecontroleerd, wisselt hier 13 – 21% van de scholen van kwalificatie. Veranderingen in keuze van covariaat leiden tot meer verschuivingen in de kwalificaties van scholen dan veranderingen in keuze van model. Dit komt overeen met het feit dat andere covariaten meer variantie kunnen verklaren dan gebeurt via alleen een andere modellering van dezelfde covariaat.

Vervolgens kunnen we meer specifiek de verschuivingen tussen verschillende modelen analyseren. We concentreren ons hierbij op de vergelijking van de beoordelingskwalificatie van scholen volgens het inspectiemodel A1 met twee meer uitgebreide correctiemodellen ABC5 en BC5. De resultaten staan in Tabel 11.

In Tabel 11 blijkt uit de vergelijking van model A1 met ABC5 dat 104 scholen 'stabiel

onder de norm' ofwel in beide modellen onder de norm scoren; 18 + 1 = 19 scholen gaan van onder de norm naar normaal/boven de norm; en 16 scholen gaan van normaal naar onder de norm. De overige scholen scoren in beide modellen steeds normaal en/of boven de norm. Anders gesteld: van 123 scholen met de beoordelingskwalificatie 'onder de norm' volgens het inspectiemodel A1, worden er 19 (15%) als 'normaal' of zelfs 'boven de norm' beoordeeld volgens model ABC5; van de 160 scholen die bij het inspectiemodel als 'normaal' worden beoordeeld, krijgen 16 (10%) de kwalificatie 'onder de norm' volgens het uitgebreidere correctiemodel ABC5. De resultaten van de vergelijking tussen A1 en BC5 lijken op die van de vergelijking A1 – ABC5: zie Tabel 11. Model BC5 heeft ten opzichte van model

Tabel 11

Vergelijking beoordeling van scholen volgens correctiemodel A1 met twee andere modellen (n scholen = 402)

Inspectiemodel A1	Model ABC5			Totaal
	onder norm	normaal	boven norm	
onder norm	104	18	1	123
normaal	16	116	28	160
boven norm	0	23	96	119
Totaal	120	157	125	402
Inspectiemodel A1	Model BC5			Totaal
	onder norm	normaal	boven norm	
onder norm	103	20	0	123
normaal	13	121	26	160
boven norm	0	21	98	119
Totaal	116	162	124	402

ABC5 als voordeel dat het eenvoudiger is (minder leerlingcovariaten bevat); bovendien lijkt BC5 minder problematisch wat betreft hoge correlaties tussen de covariaten (met name tussen opleiding ouders en leerlinggewicht bestaat een hoge correlatie).

4.4 Effecten van schoolgrootte

Naarmate een school meer leerlingen telt, ofwel groter is, zal het residu uit een multiniveau model met eenzelfde geobserveerde score groter zijn dan dat van een kleine school omdat de kleine school minder informatie (leerlingen) heeft (er is sprake van zogenaamde ‘shrunken residuals’; zie bijv. Aitken & Longford, 1986). Bij analyse van de beoordeling van schoolopbrengsten is een multiniveau model daarom te prefereren boven een plat model: een kleine school wordt minder snel ‘bij toeval’ te laag ingeschat. Bij vergelijking tussen multiniveau modellen onderling zal dit vermoedelijk weinig uitmaken omdat dan steeds sprake is van *shrinkage*. Om te verhelderen in welke mate de schoolgrootte (mede) van invloed kan zijn op de analysesresultaten in Tabel 11, hebben wij analyses uitgevoerd op de 200 grootste scholen uit het bestand. De analyses zijn uitgevoerd conform die in Tabel 11; de resultaten staan in Tabel 12.

In Tabel 12 blijkt dat er vrijwel geen verschil is bij vergelijking van model A1 (inspectiemodel) met model ABC5 (zie Tabel 11, bovenste venster). Van 64 scholen met de beoordelingskwalificatie ‘onder de norm’ volgens het inspectiemodel A1 in Tabel 12, worden er 11 (17%) als ‘normaal’ of zelfs

‘boven de norm’ beoordeeld volgens model ABC5; in Tabel 11 is dit percentage 15. Van de 77 scholen die in Tabel 12 bij het inspectiemodel als ‘normaal’ worden beoordeeld, krijgen 8 (10%) de kwalificatie ‘onder de norm’ volgens het uitgebreidere correctiemodel ABC5; in Tabel 11 is dit percentage ook 10.

5 Discussie

5.1 Samenvatting

In de huidige beoordelingssystematiek van scholen primair onderwijs gaat de inspectie met name uit van de scores op de Cito-Eindtoets Basisonderwijs van zo mogelijk alle leerlingen per school. Rekening houden met de aanvangsverschillen tussen leerlingen gebeurt via geaggregeerde schoolkenmerken (het percentage leerlingen in de schoolpopulatie met gewicht 0.3 en gewicht 1.2 zoals die door Cito bij de rapportages over de Eindtoets worden gebruikt). Het doel van ons onderzoek was te achterhalen welke bruikbare (andere) correctiefactoren zouden kunnen worden gehanteerd bij de bepaling van de eindopbrengsten van basisscholen, wat de effecten daarvan zijn, en wat de effecten zijn van correcties op verschillende niveaus.

Ter realisatie van dit doel hebben wij reeksen modellen ontworpen waarin kenmerken van leerlingen op individueel niveau en op schoolniveau zijn gebruikt om de ruwe uitkomsten op de Eindtoets Basisonderwijs te corrigeren. Gestart is met de correctie die de inspectie gebruikt. Ter uitvoering van de

Tabel 12:

Beoordeling van grootste scholen ($n = 200$) volgens vergelijking correctiemodel A1 met model ABC5

N scholen=402		Model ABC5			Totaal
Inspectiemodel A1		onder norm	normaal	boven norm	
onder norm		104	18	1	123
normaal		16	116	28	160
boven norm		0	23	96	119
Totaal		120	157	125	402
N scholen=200		Model ABC5			Totaal
Inspectiemodel A1		onder norm	normaal	boven norm	
onder norm		53	11	0	64
normaal		8	55	14	77
boven norm		0	10	49	59
Totaal		61	76	63	200

empirische analyses hebben wij gebruik gemaakt van gegevens van scholen primair onderwijs die mee hebben gedaan aan de eerste meting van het COOL-cohortonderzoek. Aanvullende kenmerken van de leerlingen uit groep 8 zijn verkregen via verschillende databestanden met registergegevens van het CBS.

Ten eerste zijn de schoolresiduen bij de verschillende correctiemodellen onderzocht. De verschillende modellering van eenzelfde leerlingcovariaat levert schoolresiduen op die maar weinig verschillen (de laagste correlatie is 0.98). Bij vergelijking van verschillende leerlingcovariaten binnen eenzelfde soort model resulteren schoolresiduen die iets meer van elkaar verschillen (de laagste correlatie is hier 0.85).

Ten tweede is nagegaan hoe groot de percentages verklaarde variantie zijn in de verschillende modellen. Bezien per afzonderlijke leerlingcovariaat, verklaren de modellen 5 (met de covariaat op individueel en geaggregeerd niveau) steeds de meeste variantie, in vergelijking met modellen die dezelfde covariaat anders modelleren. Het inspectiemodel A1 resulteert in 6.6% verklaarde variantie; bij inclusie van leerlinggewichten op individueel niveau (model A5) neemt de verklaarde variantie toe tot 11.4%. Covariaat B (opleiding ouders) voorspelt het beste de scores op de Cito-eindtoets (15.6% verklaarde variantie). Covariaat C (etnische herkomst), covariaat D (huishoudinkomen) en de drie variabelen die werden afgeleid uit het postcodegebied waarin de leerling woont (E: percentage lage inkomens; F: percentage niet-westers allochtoon; G: percentage uitkeringen) verklaren minder variantie dan de covariaten B en A. De analyses met combinaties van de covariaten A tot en met G laten soms convergentieproblemen zien. Een relatief goed model (17.1% variantie, zonder convergentieproblemen) is model ABC5 met de covariaten leerlinggewicht, opleiding ouders en etnische herkomst. Een ander, iets eenvoudiger goed model is BC5: opleiding ouders en etnische herkomst (16.6% variantie).

Een derde analyse-activiteit betrof het onderzoek van de percentages scholen die tussen modellen wisselen van beoordelingskwalificatie. In Tabel 10 wordt duidelijk dat

veranderingen in de keuze van leerlingcovariaat voor 0 – 5% van de scholen leiden tot wisseling tussen wel/niet *onder* de norm en voor 0 – 9% tot wisseling van scholen tussen wel/niet *boven* de norm. De percentages wisselingen per model 1 of 5 zijn relatief groter: veranderingen in leerlingcovariaat leiden hier voor 13 – 17% van de scholen tot wisseling tussen wel/niet *onder* de norm en voor 17 – 21% tot wisseling tussen wel/niet *boven* de norm. Dit relatief hoge percentage wisselingen wat betreft de cesuur is, gezien de werkwijze van de inspectie, mogelijk gebaseerd op slechts een klein verschil in beoordeling per school wat betreft net boven of net onder de norm. Daarnaast kan nog worden opgemerkt dat er een tendens is dat veranderingen in de keuze van leerlingcovariaat leiden tot méér verschuivingen in de kwalificaties van scholen dan veranderingen in de keuze van het model. Dit komt overeen met het feit dat andere covariaten ook meer – of ook andere – variantie verklaren dan alleen een andere modellering van dezelfde covariaat.

5.2 Beperkingen

Bij deze onderzoeksresultaten zijn wel enkele beperkende condities aanwezig. Ten eerste is voor een harde statistische toets het aantal scholen bij de overgangsanalyses van bijv. de beoordelingskwalificatie ‘onder de norm’ naar ‘normaal’, en andersom, te gering. Bovendien zijn niet-gemeten leerling- en schoolkenmerken, ‘echte’ meetfouten en (variatie in) betrouwbaarheid van de metingen ook aspecten die onze resultaten beïnvloeden. En, zoals al opgemerkt, een adequate analyse vereist in essentie een longitudinaal, multiniveau onderzoeksdesign met een hierop gebaseerde dataverzameling.

Een tweede beperking is dat onze analyses zijn uitgevoerd over één cohort leerlingen. De inspectie baseert haar oordeel echter op leerresultaten van de school wat betreft gegevens van drie schooljaren. Dit gebeurt met name om de kans op toevalsbeoordelingen te verkleinen. Mogelijk zou het percentage scholen dat in ons onderzoek van kwalificatie wisselt, lager uitvallen als we meerjarige gegevens over scholen in het onderzoek hadden kunnen betrekken.

Een derde beperking is dat wij slechts be-

paalde Cito- en andere leerlinggegevens benutten, terwijl het oordeel dat de inspectie geeft niet volledig wordt bepaald door dergelijke rekenregels. Inspecteurs kunnen bij een kwaliteitsoordeel in het toezichtkader daarvan afwijken: zij kunnen eigen wegingen bij de beoordeling van het leerlingenpubliek, de schoolomstandigheden of de schoolkwaliteit toepassen. Over de mate waarin dat voorkomt, hebben we geen informatie.

5.3 Conclusies

Uit de onderzoeksresultaten kan wat betreft bruikbare correctiefactoren het volgende worden geconcludeerd. Met name de correctiefactoren betreffende opleiding ouders (B) en etniciteit (C) voldoen, gezien de empirische resultaten, beter dan de huidige invulling van leerlinggewicht (A). In het bijzonder de modellen BC5 en ABC5 zijn relatief beter, waarbij model BC5 mogelijk voordelen heeft boven model ABC5. De argumenten zijn dat BC5 eenvoudiger is, ofwel minder variabelen bevat, en ook minder problematisch lijkt wat betreft hoge correlaties tussen de covariaten. Het voordeel van model BC5 boven het huidige inspectiemodel A1 is dat met name leerlingcovariaat B het beter doet wat betreft voorspelling dan leerlingcovariaat A. Model BC5 is dan, onder de huidige condities, het relatief beste correctiemodel.

De analyses laten daarmee zien dat verbeteringen mogelijk zijn in de wijze waarop de inspectie de leeropbrengsten van scholen beoordeelt. In de huidige werkwijze corrigeert de inspectie de ruwe resultaten op de Eindtoets Basisonderwijs voor de leerlinggewichten van de leerlingen. Deze gewichten zijn echter alleen gebaseerd op het opleidingsniveau van de ouders en differentiëren niet goed 'aan de bovenkant'. De huidige gewichtensystematiek kent alleen laag opgeleide ouders (laag = 0.3; zeer laag = 1.2; overig = 0.0). In ons onderzoek is gebleken dat, wanneer een meer uitgebreide operationalisatie van opleiding ouders wordt gebruikt, en daaraan gegevens over de etnische herkomst van de leerlingen worden toegevoegd, er duidelijk meer variantie in de leeropbrengsten van scholen wordt verklaard. Ook het niveau van operationalisering van factoren doet er toe: analyses met een model waarin ook leerling-

gegevens op individueel niveau zijn opgenomen, zijn nauwkeuriger dan analyses met variabelen op slechts schoolniveau.

Denkbare verbeteringen zijn overigens niet identiek aan direct realiseerbare verbeteringen. De inspectie beschikt op dit moment niet over informatie wat betreft het precieze opleidingsniveau van de ouders, noch over gegevens over etnische herkomst. De laatste soort gegevens komt binnenkort wel beschikbaar wanneer het onderwijsnummer voor het basisonderwijs volledig is ingevoerd, maar opleiding ouders ontbreekt dan nog. Andere factoren die bij de mogelijke optimalisering van het toezicht- en waarderingskader een rol spelen zijn: de soms optredende wisselingen in de politieke context van de beoordeling van scholen; het willen realiseren van een zo licht mogelijke bevraginglast; het gegeven dat de Cito-Eindtoets (nog) niet verplicht is; en dat de codering van het onderwijsnummer (nog) méér relevante informatie zou kunnen bevatten dan nu wordt gerealiseerd.

Het onderzoek maakt tevens duidelijk dat wanneer méér, en meer precieze, achtergrondvariabelen van leerlingen worden meegenomen in de correctie van de leeropbrengsten per school, scholen een ander oordeel van de inspectie kunnen krijgen in termen van 'onder de norm' – 'normaal' – 'boven de norm', dan in de huidige correctiesystematiek gebeurt. Dat ligt ten dele aan het niet onderscheiden van middelbare en hoge opleidingen in het huidige inspectiemodel. Het lijkt dat scholen die veel leerlingen met mbo-opgeleide ouders hebben, door de inspectie enigszins worden ondergewaardeerd. De variabele etnische herkomst speelt hierbij ook een rol. Scholen met veel autochtone achterstandsléerlingen (en weinig allochtone leerlingen) worden enigszins overgewaardeerd. Dit kan een gevolg zijn van de recente wijzigingen in de gewichtenregeling. Omdat de inspectie alleen kan corrigeren met de leerlinggewichten, volgt de wijze van corrigeren dus ook de wijzigingen in de gewichtenregeling. Scholen die voorheen veel 0.9 leerlingen hadden, dat wil zeggen leerlingen die sinds de gewichtenverandering nog maar als 0.3 of 0.0 leerlingen tellen, krijgen in de huidige systematiek minder correctie toegekend dan in het verleden.

Ondanks de beperkingen van het onderzoek geven onze analyseresultaten wel duidelijke hints voor de verantwoorde verdere ontwikkeling van de beoordelingssystematiek van scholen. Het onderwijsniveau van de ouders lijkt de belangrijkste voorspeller te zijn; etnische herkomst voegt daar nog verklarende waarde aan toe. Een multiniveau benadering via school- en leerlingkenmerken resulteert in méér voorspellende kracht dan alleen school- of leerlinggegevens. Longitudinale verzameling van leerlingkenmerken, en multiniveau evaluatie van vorderingen vanaf het schoolbegin, vormen dan de belangrijkste richtsnoeren voor een optimale schoolevaluatie én continue stimulering van leerlingen (vgl. Inspectie van het Onderwijs, 2004, 2006; Meelissen & Luyten, 2011; Mooij, 2010; Mooij & Driessen, 2008; OECD, 2008; Onderwijsraad, 2006).

Literatuur

- Aitken, M., & Longford, N. T. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistic Society*, 149, 1-43.
- Centraal Bureau voor de Statistiek. (2009). *Lage inkomens, kans op armoede en uitsluiting 2009*. Den Haag / Heerlen: Auteur.
- Centraal Bureau voor de Statistiek. (2010a). *Documentatierapport Gemeentelijke Basisadministratie (GBA) 1995-2010V1*. Den Haag/Heerlen: Auteur.
- Centraal Bureau voor de Statistiek. (2010b). *Documentatierapport Integraal-Inkomensbestand 2008V3 (Personen en Huishoudens)*. Den Haag / Heerlen: Auteur.
- Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement*, 17(3), 347-366.
- Driessen, G., Mooij, T., & Doesborgh, J. (2007). *Hoogbegaafdheid van leerlingen in het primair onderwijs: Ontwikkelingen en samenhangen met kenmerken van thuis, de groep en de school*. Nijmegen: Radboud Universiteit, ITS.
- Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J., & Van der Veen, I. (2009). *Cohortonderzoek COOL⁵⁻¹⁸. Technisch rapport basisonderwijs, eerste meting 2007/08*. Nijmegen: Radboud Universiteit, ITS / Amsterdam: SCO-Kohnstamm Instituut.
- Hox, J. J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Inspectie van het Onderwijs. (2004). *De zorg voor leerlingen met dyslexie, ADHD, autisme en hoogbegaafdheid. Een onderzoek naar de kwaliteit van handelingsplannen in het basisonderwijs in 2004*. Utrecht: Auteur.
- Inspectie van het Onderwijs. (2008). *Plusklassen, brief aan de staatssecretaris van OCW, 18 augustus 2008*. Utrecht: Auteur.
- Inspectie van het Onderwijs. (2010). *Analyse en waarderings van opbrengsten. Primair Onderwijs*. Utrecht: Auteur.
- Meelissen, M. R. M., & Luyten, H. (2011). School-effectiviteit en prestatieniveau natuuronderwijs in groep 6: Secundaire analyses op TIMMS-2007 data. *Pedagogische Studiën*, 88(5), 309-322.
- Mooij, T. (2010). Schoolontwikkeling en optimalisering van leerprocessen. In J. R. M. Gerris, J. W. Veerman, & A. Tellings (Ed.), *Jeugd- en gezinsbeleid vanuit pedagogisch perspectief. Deel 2: Uitgewerkte beleidsthema's* (pp. 249-269). Antwerpen / Apeldoorn: Garant.
- Mooij, T., & Driessen, G. (2008). Differential ability and attainment in language and arithmetic of Dutch primary school pupils. *British Journal of Educational Psychology*, 78(3), 491-506.
- Mooij, T., Smeets, E., & De Wit, W. (2011). Multi-level aspects of social cohesion of secondary schools and pupils' feelings of safety. *British Journal of Educational Psychology*, 81(3), 369-390.
- Mulder, L., Roeleveld, J., Van der Veen, I., & Vierke, H. (2005). *Onderwijsachterstanden tussen 1988 en 2002: Ontwikkelingen in basis- en voortgezet onderwijs*. Nijmegen / Amsterdam: ITS / SCO-Kohnstamm Instituut.
- Mulder, L., & Vierke, H. (2007). *Aanpassen gewichtenregeling op basis van cumulatieve gebieden*. Nijmegen: Radboud Universiteit, ITS.
- Organisation for Economic Co-operation and Development. (OECD) (2008). *Measuring Improvements in Learning Outcomes. Best Practices to Assess the Value-Added of Schools*. Retrieved from <http://www.oecd.org>

document/54/0,3746,en_2649_39263231_41701046_1_1_1_1,00.html

- Onderwijsraad. (2006). *Naar meer evidence based onderwijs. Advies*. Den Haag: Auteur.
- Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal*, 45(1), 206-230.
- Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Roeleveld, J., Driessen, G., Ledoux, G., Cuppen, J., & Meijer, J. (2011). *Doelgroep leerlingen in het basisonderwijs; Historische ontwikkeling en actuele situatie*. Amsterdam / Nijmegen: Kohnstamm Instituut / ITS.
- Roeleveld, J., Mooij, T., Fettleaer, D., & Ledoux, G. (2011). *Correctiefactoren bij opbrengstmaten in het primair onderwijs. Onderzoek ten behoeve van de Inspectie van het Onderwijs*. Amsterdam: Kohnstamm Instituut.
- Roeleveld, J., Van der Veen, I., & Ledoux, G. (2009). *Verkenning leerwinst als indicator voor onderwijskwaliteit*. Amsterdam: SCO-Kohnstamm Instituut.
- Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis. An introduction to basic and advanced multilevel modeling. 2nd edition*. London: Sage.
- Ver Eecke, E. (2004). Leerwinst als kwaliteitsindicator: Een haalbare kaart of een brug te ver? *Impuls*, 34(3), 149-163.

Noot

- 1 Met dank aan de begeleidingscommissie van de Inspectie van het Onderwijs die bij het onderzoek was betrokken

Manuscript aanvaard op: 13 juli 2012

Auteurs

Ton Mooij is manager van onderzoek en senior-onderzoeker bij het Instituut voor Toegepaste Sociale wetenschappen (ITS) van de Radboud Universiteit te Nijmegen. Tevens is hij bijzonder hoogleraar 'Onderwijstechnologie' bij het 'Centre for Learning Sciences and Technologies' (CELSTEC) van de Open Universiteit Nederland te Heerlen. **Jaap Roeleveld** is werkzaam bij het Kohnstamm Instituut van de Universiteit van Amsterdam. **Daan Fettleaer** is werkzaam bij het Instituut voor Toegepaste Sociale wetenschappen (ITS) van de Radboud Universiteit te Nijmegen. **Guuske Ledoux** is werkzaam bij het Kohnstamm Instituut van de Universiteit van Amsterdam.

Correspondentieadres: Ton Mooij, Radboud Universiteit Nijmegen, ITS, Mercator 1, Toernooiveld 212, 6525 EC Nijmegen.

Abstract

Quality judgement of primary schools: Comparing the Inspectorate's model with alternative models

To judge the quality of Dutch primary schools, the Inspectorate uses end scores of primary pupils on the national pupil monitoring system in particular. To correct for initial differences between pupils, a global estimate of parents' level of attainment in education is aggregated per school. The question for research is whether other types of correction could be used and what effects result from applying other correction types. To answer this question, seven variables at pupil level are modelled according to different random intercept multilevel regression techniques. Various types of secondary analyses are carried out on national data representing 402 schools and 8.561 pupils. The results demonstrate that two variables (detailed level of educational attainment of parents; ethnic group) produce better corrections than the correction actually used by the Inspectorate. The conclusion is that it seems possible to improve the present procedure to judge the quality of primary schools.