

# De kwaliteit van het basisonderwijs: dalen de Cito-scores?

J. Roeleveld

## Samenvatting

Terwijl ouders van leerlingen in groep 8 van het basisonderwijs steeds hoger opgeleid zijn, stijgen de gemiddelde scores van hun kinderen op de Cito-eindtoets basisonderwijs niet. Dit kan opgevat worden als een feitelijke ("virtuele") daling, die mogelijk wijst op achteruitgang van de kwaliteit van het basisonderwijs.

In dit artikel wordt nagegaan in hoeverre er over de jaren 1995, 1997, 1999 en 2001 sprake is van zo'n daling na controle voor opleidingsniveau. Gegevens uit het PRIMA-cohortonderzoek zijn geanalyseerd met een drie-niveaumodel, waarbij leerlingen worden genest binnen jaargroepen binnen scholen. De analyses wijzen op een zeer lichte tendens tot daling, die niet kan worden toegeschreven aan de stijgende deelname van scholen met veel achterstandsleerlingen aan de Cito-eindtoets.

Tegelijk blijken scores van leerlingen op andere toetsen voor prestaties en capaciteiten wel een stijgende tendens te vertonen. Daarom wordt aanbevolen om oordelen over de kwaliteit van het basisonderwijs niet in te perken tot slechts één indicator van de prestaties van leerlingen.

## 1 Achtergrond en vraagstelling

Er is de laatste jaren een groeiende belangstelling voor de output van scholen en onderwijs. Deze belangstelling komt zowel tot uiting in bijdragen vanuit wetenschappelijke hoek, zoals vorig jaar de conferentie "Het oog der natie: scholen op rapport" van de Vereniging voor Onderwijsresearch (Dijkstra, Karsten, Veenstra & Visscher, 2001) als in een stroom van bijdragen in de landelijke pers, al dan niet voorzien van lijstjes met de beste en de slechtste scholen. Ook de informatie over scholen op de website van de Inspectie van het Onderwijs is een teken van

deze publieke belangstelling voor de opbrengsten van scholen.

Bij de outputindicatoren voor het basisonderwijs neemt de Eindtoets Basisonderwijs van het Cito een vooraanstaande plaats in. Deze toets wordt op een ruime meerderheid van de Nederlandse basisscholen afgenomen in groep 8 en omvat de onderdelen Taal, Rekenen en Informatieverwerking. De scores op deze drie onderdelen worden, middels een equivaleringsprocedure, omgezet naar een over de jaren te vergelijken standaardscore op de Eindtoets (Uiterwijk & Theunissen, 2001).

Wanneer onderzoekers menen dat er iets aan de hand is met de scores op deze Cito-toets, dan is daar onmiddellijk publieke belangstelling voor. Begin vorig jaar verscheen in het dagblad "Trouw" een artikel<sup>1</sup> waarin vermeld werd dat Jungbluth op basis van gegevens uit het PRIMA-cohortonderzoek concludeert dat de resultaten op de Cito-eindtoets gedaald zijn sinds 1995. In een reactie meldde het Cito dat hun gegevens een dergelijke conclusie niet toelaten; zij vinden alleen verschillen die binnen de toevalsmarge blijven. Dronkers en Roeleveld<sup>2</sup> brachten daarop naar voren dat, als er rekening wordt gehouden met het stijgende opleidingsniveau van ouders, er inderdaad sprake is van een zekere daling van het gemiddelde op de Eindtoets Basisonderwijs. Zij spreken van een "verborgen verslechtering" van de scores.

Bij al deze beschouwingen over een (naar ook uit dit artikel zal blijken) geringe verschuiving in gemiddelde scores op een toets, die een momentopname geeft van de prestaties van leerlingen in groep 8, werd er direct in kranten en radio-uitzendingen uitgebreid gefilosofeerd over de achteruitgang van het basisonderwijs in Nederland en allerlei mogelijke oorzaken daarvoor. Als de thermometer van de Eindtoets een streepje stijgt wordt de patiënt al snel doodziek verklaard.

In dit artikel willen we nagaan of er sprake is van een daling van gemiddelde scores

op de Cito-eindtoets basisonderwijs en, indien dat zo is, of andere indicatoren van het niveau van het basisonderwijs in Nederland dan dezelfde kant opwijzen.

Centraal staan de gemiddelde scores op de Cito-eindtoets in vier jaren: 1995, 1997, 1999 en 2001<sup>3</sup>. Zoals het Cito heeft opgemerkt, verandert het landelijk gemiddelde op de Eindtoets in deze periode niet noemenswaard. Maar bij de hiervoor genoemde analyses van Jungbluth en van Dronkers en Roeleveld werd gesuggereerd dat, vanwege de stijging over de genoemde periode van het opleidingsniveau van ouders, de scores op de eindtoets naar verwachting ook hadden moeten stijgen. Uit een omvangrijke hoeveelheid onderzoeken komt het ouderlijke opleidingsniveau naar voren als een door de jaren heen onveranderlijk belangrijke determinant van schoolse prestaties van leerlingen (zie bijvoorbeeld: Dronkers, 1990; Dronkers & Ultee, 1995). Wanneer we aannemen dat het belang van dat onderwijsniveau over een tijdsspanne van zes jaar niet wezenlijk verandert, dan zouden bij een gemiddelde stijging van het opleidingsniveau ook de gemiddelden op de Cito-eindtoets moeten stijgen.

Uit de gegevens van het Cito zelf kan dit niet achterhaald worden. Alleen met behulp van databestanden waarin ook gegevens over de opleiding van de ouders zijn opgenomen, is het mogelijk deze redenering te toetsen. In dit artikel maken we daarom gebruik van gegevens uit het PRIMA-cohortonderzoek. Aan de hand daarvan zullen we allereerst nagaan of er, na controle voor opleidingsniveau en etnische achtergrond van de ouders van leerlingen, sprake is van een daling van de gemiddelde scores op de Cito-eindtoets (vraagstelling 1).

De deelname van scholen aan de eindtoets is de laatste jaren voortdurend gestegen. Mogelijk zijn er ook andere soorten scholen mee gaan doen, bijvoorbeeld meer scholen met een relatief groot aandeel kinderen uit achterstandssituaties. Daarom onderzoeken we ook of verandering van de gemiddelde Cito-score te verklaren is vanuit een verandering in het type scholen dat de Cito-toets is gaan afnemen (vraagstelling 2). Daarbij gaat het zowel om de samenstelling van de leerling-

bevolking, als om de regio en de ligging in een grote stad.

Een andere grond voor de verwachting dat de gemiddelde scores op de Cito-eindtoets hadden moeten stijgen, wordt gevonden in het feit dat er bij de toetsen uit het PRIMA-cohortonderzoek wel sprake lijkt te zijn van een stijging over de jaren, zowel bij de Taalen Rekentoetsen, als bij de scores op een test voor non-verbaal IQ. Als die toetsen een indicatie vormen van de schoolprestaties en capaciteiten van de leerlingen, dan is er dus sprake van stijgende (gemiddelde) prestaties en capaciteiten bij de leerlingen zonder de op grond daarvan te verwachten stijging van gemiddelde scores op de Cito-eindtoets. Daarom zullen we onderzoeken of er, rekening houdend met opleidingsniveau en etnische herkomst van leerlingen, sprake is van een stijging van prestaties en capaciteiten van de leerlingen, zoals die tot uitdrukking komen in de scores op andere toetsen dan de Cito-eindtoets (vraagstelling 3).

In paragraaf 2 geven we een korte beschrijving van de data waarop de analyses worden uitgevoerd en van het gehanteerde analysemodel. In paragraaf 3, 4 en 5 wordt dan achtereenvolgens ingegaan op de drie vraagstellingen. In paragraaf 6 sluiten we af met conclusies en een discussie van de uitkomsten van de analyses.

## 2 Beschrijving van data en analysemodel

Voor de analyses maken we gebruik van gegevens uit het PRIMA-cohortonderzoek. In dit onderzoek worden elke twee jaar bij een groot aantal leerlingen en scholen voor regulier en speciaal basisonderwijs gegevens verzameld. De eerste ronde van dataverzameling was in schooljaar 1994/1995 (Van Langen, Vierke & Robijns, 1996; Jungbluth, Van Langen, Peetsma & Vierke, 1996), de meest recente vond plaats in schooljaar 2000/2001 (Driessen, Van Langen & Vierke, 2002; Roeleveld & Van der Meijden, 2002). Bij elke ronde is er sprake van een referentiesteekproef, die als representatief kan worden beschouwd voor het gehele (reguliere) Nederlandse basisonderwijs.

Tabel 1

Gemiddelde scores op vier toetsen in de PRIMA-referentiesteekproeven van vier schooljaren

	94/95	96/97	98/99	00/01	SD *
Cito-eindtoets	534.5	534.2	533.9	534.3	10.0
PRIMA taalvaardigheid	1117.0	1121.9	1122.9	1119.3	36.8
PRIMA rekenvaardigheid	1196.0	1211.4	1223.8	1223.9	40.1
Nonverbaal IQ	25.5	25.8	26.2	26.4	4.3

Noot. \* Gemiddeld over de 4 schooljaren.

Wij maken gebruik van gegevens van alle leerlingen uit groep 8 van scholen uit deze referentiesteekproef van achtereenvolgens PRIMA1, PRIMA2, PRIMA3 en PRIMA4.<sup>4</sup> Bij deze leerlingen zijn onder meer toetsen voor taal en rekenen afgenomen. Bovendien zijn bij scholen die meedoen aan de Cito-eindtoets basisonderwijs de scores van dezelfde leerlingen opgevraagd.

De standaardscores op de eindtoets zijn over de jaren heen te vergelijken doordat het Cito, via een equivaleringsprocedure, de toetsresultaten van de verschillende jaren in één schaal uitdrukt.<sup>5</sup> In het PRIMA-onderzoek zijn een taal- en een rekentoets afgenomen, waarvan het resultaat wordt uitgedrukt in Taal- en Rekenvaardigheidsscores; verder is steeds een toets voor non-verbaal IQ afgenomen (Driessen, Van Langen & Oudenhoven, 1994). De taaltoets en de IQ-test voor groep 8 zijn gedurende de onderzochte jaren identiek. De Rekenvaardigheidsscores zijn in de eerste twee jaren bepaald op basis van de PRIMA-rekentoets en in de twee laatste jaren op basis van de toets Rekenen en Wiskunde uit het Cito-leerlingvolgsysteem. Hoewel de scores op deze verschillende toetsen omgezet kunnen worden in dezelfde Rekenvaardigheidsscore (Kamphuis, Mulder, Vierke, Overmaat & Koopman, 1998) moet een trend over de jaren toch met enige voorzichtigheid worden geïnterpreteerd.

Uit de schooladministraties zijn in ieder jaar gegevens verkregen over opleiding en geboorteland van de ouders. Hieruit is in PRIMA de variabele *sociaal-etnische herkomst* bepaald. Deze maakt allereerst een onderscheid naar opleiding van de ouders (hooguit lager beroepsonderwijs; middelbaar onderwijs; hoger onderwijs); alleen binnen de laagste categorie wordt vervolgens een indeling naar etniciteit gemaakt (Turks/Marok-

kaanse herkomst; overig allochtone herkomst; autochtone herkomst).

We selecteren alle leerlingen uit groep 8 van scholen uit de referentiesteekproef van PRIMA1, PRIMA2, PRIMA3 en PRIMA4. Het gaat dan om in totaal 36356 leerlingen. Daarvan is bij 22612 (62.2%) een score op de Cito-eindtoets bekend. In totaal zijn er 807 verschillende scholen bij de verschillende PRIMA-metingen betrokken; 133 daarvan hebben aan alle vier PRIMA-metingen meegedaan en 67 hebben in al die vier jaren ook aan de Cito-eindtoets meegedaan.<sup>6</sup>

In Tabel 1 geven we een overzicht van gemiddelden in de vier beschikbare schooljaren op de Cito-eindtoets, de PRIMA-Taal- en Rekenvaardigheidsscores en de scores voor non-verbaal IQ.

We zien dat de scores op de Cito-eindtoets in het tweede en derde jaar licht dalen en in het laatste jaar weer iets stijgen. De verschillen zijn echter, in overeenstemming met de populatiegegevens van het Cito, miniem. De PRIMA-taalvaardigheid neemt eerst toe, maar daalt weer wat in het laatste jaar. Bij zowel de PRIMA-rekenvaardigheid als bij het non-verbaal IQ is er sprake van een stijging.

Zoals gezegd, zijn er veranderingen in de samenstelling van de leerlingbevolking op de scholen. In Tabel 2 geven we de verdeling van de groep-8-leerlingen in de vier onderzochte schooljaren naar hun sociaal-etnische herkomst.

Het aandeel van beide groepen allochtone leerlingen van laagopgeleide ouders varieert enigszins, maar samen vormen zij steeds 9.5 tot 10% van de schoolbevolking. De groep autochtone leerlingen van laagopgeleide ouders neemt duidelijk af, terwijl er vooral bij de leerlingen van hoger opgeleide ouders sprake is van een toename. Gezien het verband tussen de ouderlijke opleiding enerzijds

Tabel 2

Verdeling van de groep 8 leerlingen naar sociaal-etnische herkomst; in de PRIMA-referentiesteekproeven van vier schooljaren (% van n)

	94/95	96/97	98/99	00/01	Verandering 95 - 01
Max Lbo; Turks/Marokkaans	5.6%	5.3%	5.1%	5.4%	- 0.2%
Max Lbo; overig allochtoon	4.4%	4.2%	4.5%	4.1%	- 0.3%
Max Lbo; autochtoon	32.4%	32.7%	28.9%	26.6%	- 5.8%
Mbo-niveau	34.9%	35.4%	36.4%	36.3%	+ 1.4%
Hbo/w.o.-niveau	22.7%	22.4%	25.0%	27.6%	+ 4.9%
n	6499	8337	8663	8612	
Onbekend	1713 (20.9%)	752 (8.3%)	945 (9.8%)	835 (8.8%)	

en schoolprestaties en schoolloopbanen anderzijds (zie paragraaf 1) is er dus alle reden om in de analyses rekening te houden met deze verschuiving in de herkomstkenmerken van de leerlingen. Tot slot laten de cijfers in de laatste rij van Tabel 2 nog zien dat van een niet onaanzienlijke groep leerlingen (vooral in het eerste onderzochte schooljaar) de herkomst onbekend is. In de analyses zal deze groep daarom als afzonderlijke categorie worden meegenomen.

In Tabel 3 worden de gemiddelde scores op de Cito-eindtoets uitgesplitst naar de sociaal-etnische herkomst van de leerlingen.

Uit Tabel 3 blijkt dat, waar de gemiddelde scores over alle leerlingen maar licht variëren, er bij de naar herkomst onderscheiden groepen grotere verschillen optreden. Bij de Turkse en Marokkaanse leerlingen van laagopgeleide ouders stijgt het gemiddelde ruim twee punten, terwijl dat bij drie andere groepen met één tot anderhalf punt daalt; de groep met hoger opgeleide ouders blijft ongeveer gelijk. De “onbekende” groep ten slotte kent ook een duidelijke stijging.

## 2.1 Analysemodel

Om na te gaan of er, wanneer we rekening houden met veranderingen in de samenstelling van de leerlingpopulatie, sprake is van een daling van de gemiddelde scores op de Cito-eindtoets maken we gebruik van multi-niveaumodellen (Goldstein, 1995; Snijders & Bosker, 1999). Hierbij wordt rekening gehouden met de afhankelijkheid in de data door de nesting van leerlingen binnen scholen. Naast deze meer “traditionele” nesting van leerlingen binnen scholen hebben we hier ook te maken met een andere nesting. Om na te gaan of er sprake is van een daling van scores op de Cito-eindtoets onderzoeken we binnen de scholen leerlingen uit groep 8 in de schooljaren 94/95, 96/97, 98/99 en 2000/01. Daardoor krijgt de nesting een drie-niveaustructuur: leerlingen binnen (school-) jaren binnen scholen.

We maken gebruik van gegevens van zoveel mogelijk leerlingen. Ook wanneer er bij scholen gegevens over bepaalde jaren ontbreken, bijvoorbeeld doordat de school toen niet aan PRIMA- en/of de Cito-eindtoets heeft

Tabel 3

Gemiddelde scores op de Cito-eindtoets naar sociaal-etnische herkomst; in de PRIMA-referentiesteekproeven van vier schooljaren

	94/95	96/97	98/99	00/01	Verandering 95 - 01
Max Lbo; Turks/Marokkaans	524.1	526.1	525.8	526.2	+ 2.2
Max Lbo; overig allochtoon	529.6	527.8	529.4	528.8	- 0.8
Max Lbo; autochtoon	531.8	530.9	530.4	530.7	- 1.2
Mbo-niveau	536.5	535.8	534.9	535.0	- 1.5
Hbo/w.o.-niveau	539.7	539.3	538.8	539.6	- 0.1
Onbekend	532.8	535.6	535.2	534.6	+ 1.7

meegedaan, kunnen de gegevens over de andere meetjaren wel in de analyses worden opgenomen.

In het model is de *score op de Cito-eindtoets* (en in paragraaf 5 op de PRIMA-toetsen) de afhankelijke variabele. Om de veranderingen in gemiddelde over de jaren te modelleren, wordt er voor de jaren 1997, 1999 en 2001 een dummyvariabele in het model opgenomen; het eerste jaar 1995 is daarbij de referentie. Veranderingen in de samenstelling van de leerlingbevolking worden gemodelleerd door de sociaal-etnische herkomst van de leerlingen in het model op te nemen. Dit gebeurt in de vorm van dummyvariabelen met de categorie *middelbaar opgeleide ouders* als referentie. Bovendien kunnen interactietermen tussen herkomst en meetjaar worden opgenomen.

De verandering in kenmerken van scholen, waarvan Cito-scores bekend zijn, worden in de analyse betrokken door kenmerken van deze scholen in het model op te nemen. Dit komt nader aan de orde in paragraaf 4.

De parameters van alle modellen zijn geschat met het programma MLwiN (Rasbash, et al., 2000).

### 3 Verandering van Cito-scores

Onze eerste vraag is of er sprake is van daling van gemiddelde Cito-scores, wanneer we rekening houden met veranderingen in de herkomst van de leerlingen. Ter beantwoording daarvan hanteren we in de analyse de volgende opbouw van modellen:

- 0 we beginnen met een nulmodel, waarin geen verklarende factoren zijn opgenomen; dit model geeft alleen de variantie in de score op de Cito-eindtoets, gesplitst over de drie niveaus: leerling, jaar en school;
- 1a dan onderzoeken we het “kale” effect van drie dummyvariabelen die de vier onderzochte jaren weergeven; daaruit leiden we verschillen tussen de jaren af, zonder dat rekening gehouden wordt met kenmerken van de leerlingen (en eventuele verschuivingen daarin);
- 1b vervolgens onderzoeken we een model met alleen de sociaal-etnische herkomst

van de leerlingen; ook de leerlingen met onbekende herkomst worden d.m.v. een dummyvariabele in deze analyse meegenomen;

- 2 het volgende model combineert modellen 1a en 1b: zowel de sociaal-etnische herkomst als de jaarvariabelen worden opgenomen; door vergelijking met de eerdere modellen wordt duidelijk of er verschillen tussen de jaren blijven bestaan wanneer we rekening houden met de herkomst van de 8<sup>e</sup>-groepers uit de verschillende jaren;
- 3 ten slotte nemen we interactietermen op tussen het jaar en de sociaal-etnische herkomst van de leerlingen; in dit model kan het effect van sociaal-etnische herkomst in het ene schooljaar anders (sterker of zwakker) zijn dan in het andere jaar.

Tussen opeenvolgende modellen zullen we ook steeds de verbetering van de ‘fit’ bij de data bepalen.

In Tabel 4 geven we een overzicht van de parameterschattingen. Bij de jaareffecten worden tussen haakjes de standaardfouten vermeld, om de significantie van deze effecten te kunnen beoordelen. Bij de andere onafhankelijke variabelen laten we deze omwille van de ruimte weg (alle herkomsteffecten zijn significant; significante interacties zijn vet gedrukt).

We zullen de opeenvolgende modellen en de conclusies daaruit kort bespreken. Het *nulmodel* geeft, als gezegd, alleen een splitsing van de variantie over de drie niveaus (83% op individueel niveau, 4% op jaar-niveau en 13% op schoolniveau).

In Model 1a wordt het verschil tussen de jaren geschat. Voor geen van de jaarvariabelen wordt een significant effect gevonden en de toevoeging van de jaarvariabelen levert geen significante verbetering van de fit van het model bij de data op. De gemiddelde scores op de Cito-eindtoets verschillen dus niet wezenlijk tussen deze vier meetjaren.

Model 1b geeft de effecten van de sociaal-etnische herkomst op de Cito-eindtoets. Al deze effecten zijn significant. Vergeleken met de referentiecategorie (leerlingen met ouders met tenminste een opleiding op Mbo-niveau) behalen leerlingen met lager opgeleide ouders, zowel van Turks/Marokkaanse, als van

Tabel 4

Parameterschattingen van opeenvolgende multiniveaumodellen voor Cito-eindscore; alle scholen en leerlingen waarover Cito-gegevens beschikbaar zijn in enig jaar ( $n = 22612$ )

Model	0	1a	1b	2	3	4
Intercept	533.9	533.9	535.2	535.5	536.0	535.8
Jaar (1995 referentie)*						
1997		-0.06 (.28)		-0.17 (.27)	-0.50 (.38)	-0.52 (.28)
1999		-0.35 (.30)		<b>-0.76 (.28)</b>	<b>-1.37 (.38)</b>	<b>-1.14 (.30)</b>
2001		0.23 (.31)		-0.31 (.30)	<b>-1.21 (.39)</b>	<b>-0.78 (.30)</b>
Herkomst (Mbo referentie)**						
Turks/Marok.			-8.01	-8.02	-10.49	-10.30
Overig allocht.			-5.51	-5.50	-5.76	-5.52
Autocht Lbo			-4.45	-4.45	-4.55	-4.46
Hbo			3.74	3.74	3.07	3.75
Onbekend			-2.51	-2.56	-4.25	-4.05
Interacties herkomst met jaar *						
Tuma97					<b>2.74</b>	<b>2.82</b>
Tuma99					<b>2.75</b>	<b>2.51</b>
Tuma01					<b>3.31</b>	<b>2.81</b>
Ovall97					-0.99	
Ovall99					0.80	
Ovall01					0.87	
Lbo97					-0.15	
Lbo99					0.04	
Lbo01					0.41	
Hbo97					0.32	
Hbo99					0.77	
Hbo01					<b>1.32</b>	
Onbek97					1.83	1.87
Onbek99					<b>2.77</b>	<b>2.52</b>
Onbek01					<b>3.57</b>	<b>3.05</b>
Varianties						
School	13.39 (1.16)	13.42 (1.16)	7.94 (.82)	7.94 (.81)	7.82 (.80)	7.83 (.80)
Jaar	3.96 (.51)	3.89 (.51)	4.09 (.49)	3.99 (.49)	3.95 (.49)	3.98 (.49)
Leerling	85.03 (.82)	85.03 (.82)	76.40 (.74)	76.40 (.74)	76.28 (.74)	76.32 (.73)
Verklaarde variantie	-	0.0%	13.6%	13.7%	14.0%	13.9%
Verbetering fit *	-	4.3	<b>2528.4</b>	9.3	<b>42.2</b>	-12.8
T.o.v. Model (aantal vrijheidsgraden)		0 (3 df)	0 (5 df)	1b (3 df)	2 (10 df)	3 (9 df)

Noot. \* Vet gedrukt indien  $p < .01$ . \*\* Alle herkomsteffecten zijn significant.

overig allochtone, als van autochtone herkomst, lagere scores op de Cito-eindtoets, terwijl leerlingen met ouders met minstens Hbo-niveau hogere scores behalen. Leerlingen waarvan de herkomst onbekend is behalen lagere scores.

In Model 2 worden de twee vorige modellen gecombineerd, en vergeleken met Model 1b levert dit model een significant betere fit op ( $p < .03$ ). Dit betekent dat er, als we rekening houden met de herkomst van de leerlingen, sprake is van systematische verschillen

tussen de schooljaren. De effecten van de jaren worden nu alle drie negatief geschat. Voor 1997 is het effect klein en niet significant. Voor het volgende jaar 1999 is het effect sterker en significant, terwijl het voor jaar 2001 weer kleiner en niet significant is. De schattingen van de herkomsteffecten verschillen vrijwel niet van die in Model 1b.

In Model 2 gaan we er vanuit dat de effecten van sociaal-etnische herkomst gelijk zijn over de vier onderzochte schooljaren. In Model 3 laten we deze veronderstelling los

en worden eventuele verschillen in effecten gemodelleerd met interactietermen. De parameterschattingen laten zien dat het effect van laagopgeleide Turks/Marokkaanse herkomst en het effect van onbekende herkomst over de jaren heen significant anders is: in 1995 is het effect meer negatief dan in de drie volgende onderzochte schooljaren, die daardoor positieve interactietermen met de twee herkomstvariabelen hebben. Dit komt overeen met het beeld uit Tabel 3 waarin deze beide groepen een stijging vertonen, tegenover de dalende tendens bij de overige groepen (en in het bijzonder bij de groep leerlingen van middelbaar opgeleide ouders, die in deze analyse als referentiecategorie wordt gebruikt). De interpretatie is dat de Turks/Marokkaanse leerlingen wat zijn ingelopen op de kinderen van ouders met Mbo-niveau (vergelijk Tabel 3). En blijkbaar komen de leerlingen waarvan we de herkomst niet kennen in 1995 meer uit laag scorende herkomstgroepen en zijn ze in de drie volgende jaren meer gespreid naar werkelijke herkomst.

De effecten van de jaarvariabelen worden duidelijk sterker (negatief) in dit model, vergeleken met Model 2. Voor het jaar 1997 zijn ze niet significant, voor beide volgende jaren wordt geschat dat leerlingen respectievelijk 1.4 en 1.2 punten lager scoren dan in 1995. Bij een standaarddeviatie van de Cito-scores van 10 gaat het dan om effectgroottes in de orde van 0.14 en 0.12.

Met betrekking tot de eerste vraagstelling kunnen we uit Model 3 concluderen dat als we rekening houden met de herkomst van leerlingen in groep 8, met verschuivingen daarin en met over de jaren wisselende samenhangen (interacties) van die herkomst met de score op de eindtoets, er sprake is van een daling van de score op de Cito-eindtoets tussen de afname in 1995 en die in 2001. De daling kondigt zich bij de afname in 1997 al aan, maar was toen nog niet groot genoeg om statistisch significant te zijn. De daling die in 1999 gevonden wordt zet zich niet door in 2001 en het verschil met 1995 wordt iets minder groot.

Op grond van Model 3 kunnen we een optimaal model bepalen dat de basis vormt voor de volgende analyse in dit artikel. In dit model worden alleen de interactietermen van

de twee herkomstcategorieën *Turks/Marokkaans* en *onbekend* opgenomen.<sup>7</sup> De parameterschattingen voor dit “optimale” Model 4 staan in de laatste kolom van Tabel 4. In de laatste regel van de tabel zien we dat de “verslechtering” van de fit van het model bij de data niet significant is. Anders gezegd: de data worden evengoed beschreven door het zuiniger optimale Model 4 als door het uitgebreide Model 3.

#### 4 Veranderingen in kenmerken van deelnemende scholen

We richten ons in deze paragraaf op de tweede vraagstelling: de mogelijke verklaring van de daling van de gemiddelde scores op de Cito-eindtoets door kenmerken van de scholen. Uit verkennende analyses (Vierke & Mulder, 2001) is duidelijk geworden dat er zich in de loop van de hier onderzochte periode een zekere verschuiving heeft voorgedaan in het type scholen in de PRIMA-referentiesteekproeven, waarop de Cito-eindtoets is afgenomen. In het bijzonder is er sprake van een toename van scholen in de grote en middelgrote steden en, vermoedelijk samenhangend, van scholen met een groot aandeel allochtone leerlingen.

Het ligt voor de hand om te veronderstellen dat de hiervoor geconstateerde daling van de scores op de Cito-eindtoets in elk geval ten dele is toe te schrijven aan deze veranderingen in kenmerken van de onderzochte scholen. In deze paragraaf willen we deze hypothese toetsen door enkele kenmerken van de scholen in het multiniveaumodel op te nemen en vervolgens na te gaan of het significante “jaareffect” daardoor verdwijnt. We hebben daartoe voor elke school in elk van de onderzochte jaren bepaald:<sup>8</sup>

- de schoolscore (SE-score) op basis van gegevens van het Ministerie van OCenW; deze wordt bepaald aan de hand van de leerlinggewichten: hoe hoger de SE-score, des te groter het aandeel leerlingen met een gewicht van 1.25 en/of 1.90.
- de ligging in een grote of middelgrote stad (een dummyvariabele);
- het percentage leerlingen met laagopgeleide ouder van Turkse of Marokkaanse

- herkomst; op basis van PRIMA-gegevens afkomstig uit schooladministraties;
- het percentage leerlingen met laagopgeleide ouders van overige allochtone herkomst; opnieuw op basis van PRIMA-gegevens<sup>9</sup>.

Daarnaast kwam uit de analyses van Jungbluth<sup>10</sup> naar voren dat er mogelijk sprake is van relevante verschillen tussen regio's. Om daar enige indruk van te krijgen hebben we de provincies van Nederland ingedeeld in de regio's Noord-Oost, Midden-West en Zuid<sup>11</sup>. Ook de *regio van de school* wordt nu als variabele in de analyse opgenomen.

We hanteren opnieuw een opbouw van modellen voor de analyse:

- 4 we herhalen het "optimale" model uit de vorige paragraaf;
- 5 in de volgende modellen nemen we één voor één de kenmerken van de scholen in de verschillende jaren op; de beide percentages leerlingen van onderscheiden allochtone herkomst combineren we in één Model 5c;
- 6 combinatie van de vorige modellen: alle schoolkenmerken bij elkaar.

In Tabel 5 worden de uitkomsten van de analyses gepresenteerd. Bij de effecten van de jaarvariabelen en de schoolkenmerken worden tussen haakjes de standaardfouten vermeld.

We bespreken weer stapsgewijs de con-

Tabel 5

*Parameterschattingen van modellen voor Cito-eindscore met schoolkenmerken*

Model	4	5a	5b	5c	5d	6	7
Intercept	535.8	536.2	542.2	536.2	535.7	539.2	537.0
Jaar (1995 referentie)*							
1997	-0.52 (.28)	-0.50 (.28)	-0.82 (.29)	-0.52 (.28)	-0.51 (.29)	-0.63 (.29)	-0.51 (.28)
1999	-1.14 (.30)	-1.22 (.30)	-1.33 (.30)	-1.09 (.30)	-1.14 (.30)	-1.27 (.30)	-1.4 (.29)
2001	-0.78 (.30)	-0.77 (.30)	-0.76 (.30)	-0.69 (.30)	-0.72 (.30)	-0.78 (.30)	-0.79 (.30)
Herkomst (Mbo referentie)**							
Turks/Marok.	-10.30	-10.13	-9.83	-9.68	-10.40	-9.67	-9.68
Overig allocht.	-5.52	-5.43	-5.30	-5.29	-5.56	-5.27	-5.33
Autocht Ibo	-4.46	-4.46	-4.44	-4.45	-4.45	-4.43	-4.43
Hbo	3.75	3.76	3.74	3.74	3.74	3.73	3.74
Onbekend	-4.05	-3.99	-3.98	-3.77	-4.12	-3.90	-3.89
Interacties herkomst met jaar *							
Tuma97	2.82	2.77	2.64	2.74	2.83	2.67	2.70
Tuma99	2.51	2.44	2.41	2.33	2.52	2.34	2.31
Tuma01	2.81	2.75	2.85	2.63	2.88	2.77	2.67
Onbek97	1.87	1.84	1.87	1.67	1.92	1.79	1.76
Onbek99	2.52	2.50	2.53	2.35	2.58	2.49	2.46
Onbek01	3.05	2.98	2.93	2.74	3.14	2.83	2.78
Schoolkenm. per jaar (Middel)grote stad		-1.28 (.30)				-0.92 (.31)	-1.12 (.30)
Schoolscore (100-180)			-0.058 (.010)			-0.025 (.015)	
% Turks/Marok. leerlingen (0-100)				-0.057 (.012)		-0.036 (.015)	-0.058 (.011)
% overig allochtone leerlingen (0-100)				-0.028 (.022)		0.009 (.023)	
Regio Noord-Oost					-1.29 (.40)	-1.83 (.39)	-2.27 (.34)
Regio Zuid					1.09 (.34)	0.71 (.33)	
Varianties							
School	7.83 (.80)	7.50 (.78)	6.87 (.75)	6.99 (.76)	6.78 (.74)	5.82 (.69)	6.08 (.70)
Jaar	3.98 (.49)	3.96 (.48)	4.11 (.49)	4.03 (.48)	4.08 (.49)	4.06 (.49)	4.02 (.49)
Leerling	76.32 (.73)	76.31 (.74)	76.31 (.74)	76.32 (.74)	76.32 (.74)	76.31 (.74)	76.30 (.73)
Verklaarde variantie	13.9%	14.3%	14.7%	14.7%	14.8%	15.8%	15.6%
Verbetering fit *		18.0	33.0	33.0	39.9	92.2	83.1
T.o.v. Model (aantal vrijh. graden)		4 (1 df)	4 (1 df)	4 (2 df)	4 (2 df)	4 (6 df)	4 (3 df)

Noot. \* *Vet* gedrukt indien  $p < .01$ . \*\* Alle herkomsteffecten zijn significant.



clusies die uit de analyses met de opeenvolgende modellen kunnen worden getrokken. De eerste drie modellen met schoolkenmerken (5a, 5b en 5c) laten de verwachte effecten zien:

- In de (middel)grote steden is de Cito-score lager (gemiddeld 1.28) dan in de rest van Nederland.
- Hoe hoger de SE-score van de school, des te lager de Cito-score.
- Hoe hoger het percentage allochtone leerlingen, hoe lager de Cito-score; (bij het percentage Turks/Marokkaanse leerlingen is het effect significant, bij het percentage overig allochtone leerlingen niet).

Let wel: al deze effecten vinden we na controle voor de individuele herkomst van de leerlingen.

Het model waarin de regio van de school is opgenomen (Model 5d) laat zien dat in het Noord-Oosten van het land gemiddeld ruim één punt lager wordt gescoord op de Cito-eindtoets dan in het Midden-Westen (de referentiecategorie); in het Zuiden van het land liggen de gemiddelde scores juist een punt hoger.

Een aantal van de schoolkenmerken hangen onderling nogal samen en de combinatie in Model 6 levert dan ook kleinere effect-schattingen op voor de meeste variabelen dan in de afzonderlijke modellen. In een zuiniger Model 7 zijn daarom alleen de significante schoolkenmerken geselecteerd, met negatieve effecten voor (middel)grote steden, het percentage Turks/Marokkaanse leerlingen en de regio Noord-Oost. Uit extra analyses blijkt dat interacties van deze schoolkenmerken met jaar niet significant zijn en ook geen beter fittend model opleveren (niet in tabel weergegeven; zie Roeleveld, 2001). De effecten zijn dus van jaar tot jaar gelijk.

Elk van de schoolkenmerken afzonderlijk (Model 5a t/m 5d) en hun combinatie in Model 6 en Model 7 verbeteren de fit en de verklarende kracht van het model vergeleken met Model 4 zonder schoolkenmerken. Het zijn daarmee relevante variabelen ter verklaring van de variantie in de scores op de Cito-eindtoets. Maar voor onze vraagstelling is het vooral van belang te kijken naar het effect van de jaarvariabelen. De negatieve tendens in deze variabelen (duidend op een daling

van de Cito-score) verdwijnt zeker niet door de toevoeging van de schoolkenmerken in het model. We komen daarom tot de conclusie dat de daling van de Cito-scores, die we in de vorige paragraaf constateerden, niet terug te voeren is op het in sterkere mate deel gaan nemen van scholen uit de steden en/of scholen met veel allochtone leerlingen.

## 5 Andere indicatoren voor prestaties van leerlingen in groep 8

De redenering bij de voorgaande analyses was dat we weliswaar geen opvallende verschillen zien in de gemiddelde scores op de Cito-eindtoets over een reeks van jaren, maar dat we dergelijke verschillen eigenlijk wel hadden kunnen verwachten. De leerlingen in groep 8 hebben in toenemende mate ouders met middelbare en hogere opleidingen en zijn daarmee van huis uit beter toegerust voor school. We hadden verwacht dat zich dat zou uitdrukken in een stijging van de gemiddelde scores op de eindtoets, maar deze is uitgebleven.

Er is nog een andere reden waarom we een stijging van de gemiddelde scores op de Cito-eindtoets hadden verwacht. Dezelfde leerlingen die de eindtoets afleggen hebben in ongeveer dezelfde periode<sup>12</sup> in het kader van het PRIMA-onderzoek ook enkele andere toetsen afgelegd en op die toetsen zien we in grote lijnen wel een stijging van de gemiddelden (zie Tabel 1). Wanneer we de scores op deze PRIMA-toetsen beschouwen als een meting van de schoolprestaties (Taal en Rekenen) en de capaciteiten (IQ-scores) van de leerlingen in groep 8, dan is er dus sprake van een stijging van die prestaties en capaciteiten en zouden we op grond daarvan mogen verwachten dat ook de scores op de Cito-eindtoets stijgen.

In deze laatste paragraaf willen we daarom nagaan of er inderdaad sprake is van een (significante) stijging van gemiddelde scores op deze drie in het PRIMA-onderzoek afgenomen toetsen, wanneer we ook hierbij gebruik maken van multiniveaumodellen waarin zowel rekening kan worden gehouden met de geneste structuur van de gegevens als met het stijgende opleidingsniveau van de ouders.

We hebben daartoe voor dezelfde groep leerlingen waarover we in de paragrafen 3 en 4 gegevens hebben gepresenteerd, nu analyses uitgevoerd met als afhankelijke variabelen respectievelijk de *PRIMA-taalvaardigheid en -rekenvaardigheid* en de *score op de toets voor non-verbale intelligentie*. In Tabel 6 geven we voor elke afhankelijke variabele afzonderlijk een model, waaruit alleen de verandering in gemiddelde over de jaren blijkt en een optimaal model, waarin gecontroleerd wordt voor de herkomst van de leerlingen en

waarin alle significante interacties van herkomst met jaar zijn opgenomen.

De Modellen 1 laten voor alle drie toetsen een stijging ten opzichte van het referentiejaar 1995 zien. Bij de taaltoets is de stijging in het laatste jaar echter minder en het gemiddelde in dat jaar verschilt niet significant van het gemiddelde in 1995. Bij de IQ-scores is de stijging in het tweede onderzoeksjaar nog te gering om significant te zijn.

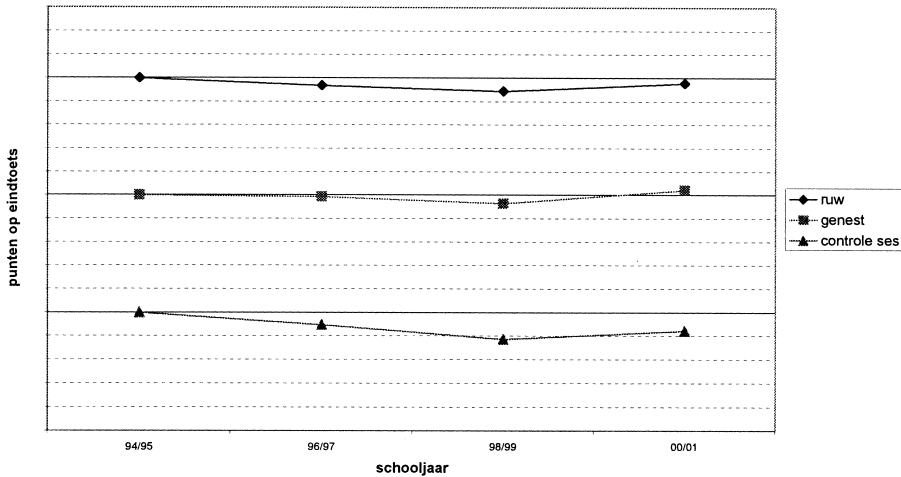
In de optimale modellen zien we steeds de verwachte effecten van de sociaal-etnische

Tabel 6

Parameterschattingen van opeenvolgende multiniveaumodellen voor drie PRIMA-toetsen

Model	Taal		Rekenen		IQ	
	Model 1	Optimaal	Model 1	Optimaal	Model 1	Optimaal
Intercept	1118.1	1122.5	1197.4	1204.4	25.6	25.9
Jaar (1995 referentie)*						
1997	4.7 (.9)	4.5 (.9)	16.3 (1.2)	14.8 (1.2)	0.28 (.12)	0.25 (.11)
1999	4.9 (1.0)	4.1 (.9)	26.3 (1.2)	21.4 (1.3)	0.65 (.12)	0.54 (.12)
2001	2.0 (1.0)	1.0 (.9)	27.3 (1.3)	21.5 (1.4)	0.90 (.12)	0.74 (.12)
Herkomst (Mbo referentie)**						
Turks/Marok.		-33.4		-33.6		-1.75
Overig allocht.		-21.9		-15.4		-1.17
Autocht lbo		-11.5		-16.8		-1.02
Hbo		11.6		10.8		1.00
Onbekend		-5.7		-13.5		-0.61
Interacties herkomst met jaar**						
Tuma97				11.5		
Tuma99				17.3		
Tuma01				19.4		
Ovall97						
Ovall99						
Ovall01						
Lbo97						
Lbo99				6.6		
Lbo01				7.8		
Hbo97						
Hbo99						
Hbo01						
Onbek97				6.3		
Onbek99				9.5		
Onbek01				12.8		
Varianties						
School	134.3	55.6	130.8	98.9	0.67	0.39
Jaar	25.6	27.4	80.8	79.7	0.57	0.55
Leerling	1200.2	1119.1	1302.6	1228.7	17.56	17.09
Verklaarde variantie	0.2%	11.8%	7.2%	13.8%	0.5%	4.5%

Noot. \* Vet gedrukt  $p < .01$ . \*\* Alleen significante effecten zijn opgenomen.



Figuur 1. Verandering in de gemiddelde scores op de Cito-eindtoets vanaf 1995; in verschillende modelvarianten.

herkomst van de leerlingen. Bij taal en IQ is geen enkele van de interactie-effecten significant; bij rekenen is dat wel het geval. Net als eerder bij de scores op de Cito-eindtoets zien we bij rekenen dat leerlingen van Turks/Marokkaanse herkomst een deel van hun achterstand op de leerlingen uit de referentiegroep inlopen; hetzelfde geldt voor de leerlingen waarvan de herkomst onbekend is. Ook bij de autochtone leerlingen met laagopgeleide ouders is er in de laatste twee onderzoeksjaren sprake van het inlopen van de achterstand op de referentiegroep.

De tendens tot stijging van de gemiddelden, met als uitzondering taal in het laatste onderzoeksjaar, blijft ook in de optimale modellen bestaan. De omvang van de stijging is in deze modellen wel wat kleiner dan in de modellen waarin geen rekening wordt gehouden met de herkomst van de leerlingen. Een deel van de stijgende tendens is dus toe te schrijven aan het stijgende opleidingsniveau van de ouders van de leerlingen.

Onze conclusie is dat er zowel bij de prestaties van leerlingen, vooral bij rekenen maar ook wel bij taal, als bij hun capaciteiten, zoals gemeten door een toets voor non-verbale IQ, sprake is van een stijging van de gemiddelden tussen 1995 en 2001. Zoals we eerder hebben gezien komt deze stijging niet tot uitdrukking in een stijging van de gemiddelden op de Cito-eindtoets.

## 6 Conclusies en discussie

In deze paragraaf zullen we puntsgewijs de belangrijkste conclusies uit de analyses in dit artikel nog eens samenvatten. Enkele trends in deze conclusies zullen we proberen grafisch te illustreren.

Over de vier onderzochte schooljaren verandert er op zichzelf weinig in de gemiddelde score op de Cito-eindtoets basisonderwijs. De zeer lichte daling over de jaren 95-97-99 zet niet door in schooljaar 2000/01. Als we 1995 op 0 stellen, dan zijn de gemiddelden in de volgende jaren (Tabel 1): -0.3 ; -0.6 ; -0.2 (zie bovenste lijn in Figuur 1)

Gezien de geneste structuur van de data is multiniveau-analyse het meest geëigend. Met een dergelijke modellering, zonder dat er nog voor andere variabelen wordt gecontroleerd, vinden we voor de ontwikkeling in gemiddelden over vier schooljaren (Model 1a in Tabel 4): -0.1 ; -0.4 ; 0.2 (zie middelste lijn in Figuur 1)

Zoals gezegd komen er meer leerlingen met middelbaar en hoog opgeleide ouders en minder met laagopgeleide ouders. Op grond daarvan zouden we mogen verwachten dat de gemiddelden op de Cito-eindtoets zouden stijgen. Wanneer we deze verwachting modelleren door te controleren voor de sociaal-etnische herkomst van de leerlingen, dan vinden we voor de ontwikkeling in gemiddelden

over vier schooljaren (optimale Model 4, Tabel 4): -0.5 ; -1.1 ; - 0.8 (derde lijn in Figuur 1).

De poging om de dalende tendens te verklaren met een aantal kenmerken van die (wisselende) scholen is weinig succesvol gebleken. Het model waarin enkele relevante schoolkenmerken zijn opgenomen, schat de daling t.o.v. 1995 als (Model 7 in Tabel 5): -0.5 ; -1.2 ; -0.8.

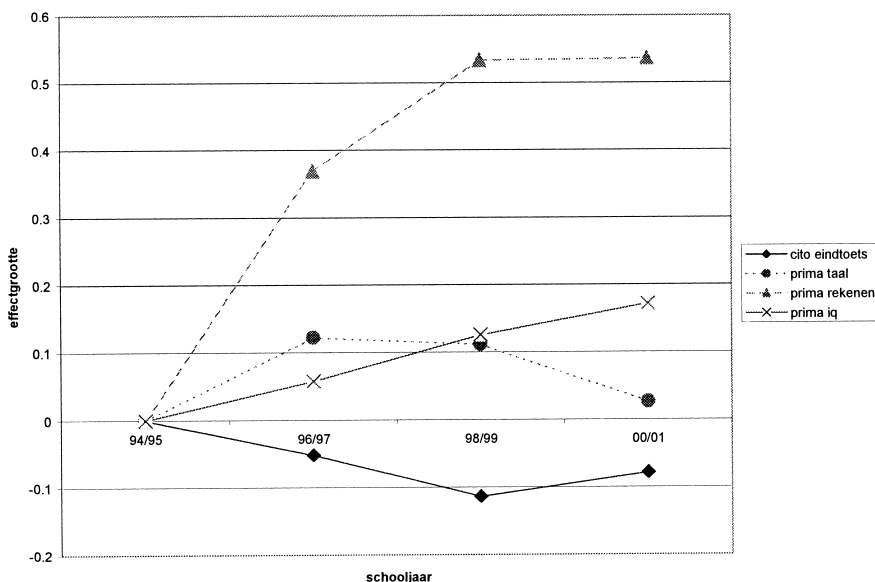
De omvang van de daling van de scores is bij alle gehanteerde modellen bescheiden. Bij een standaarddeviatie van de Cito-eindtoets van ongeveer 10 punten kan de daling in termen van effectgrootte geschat worden als hooguit 0.12 van een standaarddeviatie op de Cito-eindtoets.<sup>13</sup>

In een recent artikel vond Webbink (2002) geen daling van Cito-scores na controle voor veranderingen in kenmerken van de leerlingpopulatie. Webbink heeft voor zijn analyse echter ook scholen uit de aanvullende steekproef van PRIMA gebruikt. Hierdoor zijn bij hem scholen met een hoog aandeel Turks/Marokkaanse leerlingen oververtegenwoordigd. Uit paragraaf 3 bleek dat juist deze leerlingen een deel van hun achterstand op de referentiegroep (leerlingen met middelbaar opgeleide ouders) inlopen. Dit kan de in dit artikel gesignaleerde dalende tendens ge-

compenseerd hebben, zodat Webbink deze tendens niet vindt.

De variabele *opleiding van de ouders* kan gezien worden als een 'proxy' voor een thuis-klimaat dat bevorderend is voor het behalen van goede onderwijsresultaten. De vooronderstelling bij de analyses in dit artikel is dat de betekenis van deze variabele over de jaren heen niet verandert. Dat lijkt ook onwaarschijnlijk over een tijdspanne van slechts zes jaar, terwijl er (behalve voor Turks/Marokkaanse leerlingen) ook geen significante interactie-effecten werden gevonden die wijzen op een verschuiving in betekenis. De verwachting dat bij een stijging van het opleidingsniveau van nieuwe generaties ouders ook de scores op de Eindtoets hadden moeten stijgen, lijkt ons dan ook plausibel.

Een probleem bij dit alles is dat waar het gemiddelde op de Cito-eindtoets bij de verschillende analyses licht daalt, de scores op enkele PRIMA-toetsen juist een stijgende tendens vertonen. Wanneer we die toetsen beschouwen als een indicatie van de schoolprestaties (taal en rekenen) en de capaciteiten (IQ) van leerlingen in groep 8, dan komt deze stijging van prestaties en capaciteiten dus niet tot uitdrukking in de scores op de Cito-eindtoets.



Figuur 2. Verandering in de gemiddelde scores op de Cito-eindtoets en drie PRIMA-toetsen vanaf 1995; in optimale modellen met controle voor de herkomst van de leerlingen.

In Figuur 2 illustreren we dit verschil in trends voor de verschillende toetsen. We gebruiken daarbij de geschatte effecten per jaar uit de verschillende optimale modellen, die steeds een multiniveaustruktuur hebben en waarin gecontroleerd wordt voor de sociaal-etnische herkomst van de leerlingen (voor de Cito-eindtoets Model 4 uit Tabel 4; voor de PRIMA-toetsen de optimale modellen uit Tabel 6). Omwille van de onderlinge vergelijking zijn alle jaareffecten uitgedrukt in effectgroottes: het effect gedeeld door de standaarddeviatie van de betreffende variabele.

Figuur 2 illustreert de discrepantie die er bestaat tussen de trends bij de verschillende toetsen:

- De gemiddelde scores op de Cito-eindtoets dalen eerst licht en stijgen weer iets in het laatste jaar.
- De gemiddelde scores op de PRIMA-taaltoets stijgen eerst licht en dalen weer in het laatste jaar.
- De gemiddelde scores op de IQ-test uit PRIMA vertonen een continu stijgende lijn; de laatste score is 0.17 standaarddeviatie hoger dan de score uit het eerste jaar.
- De gemiddelde scores op de PRIMA-rekentoets stijgen eerst duidelijk en blijven in het laatste jaar gelijk; bij deze scores merken we opnieuw op dat in de laatste twee jaar een andere toets is afgenomen dan in de eerste twee jaar, zodat deze stijgende trend met enige voorzichtigheid moet worden bezien.

Het verschil in trends tussen de diverse toetsen kan verschillende oorzaken hebben. Zoals gezegd, is de PRIMA-rekentoets in de afgelopen jaren gewijzigd, maar de PRIMA-taaltoets en de testreeksen (IQ-toetsen) zijn identiek gebleven over de vier achtereenvolgende PRIMA-metingen. Het is zeer onwaarschijnlijk dat stijgende scores op die toetsen komen doordat scholen hun leerlingen speciaal trainen voor die toetsen (*teaching to the test*) of doordat leerlingen meer geholpen worden bij de afnames. We mogen aannemen dat de stijging wel degelijk wijst op toenemende kennis en vaardigheden bij de leerlingen.

De Cito-eindtoets is echter elk jaar anders en de daar uitblijvende stijging zou te maken

kunnen hebben met onvolkomenheden in de equivaleringsprocedure, die de scores op deze wisselende toetsen over de jaren heen vergelijkbaar moeten maken. Aan de andere kant is van de PRIMA-taaltoets bekend dat deze meer geschikt is voor de onderkant van de prestatieverdeling (Driessen, e.a., 1994), terwijl de Cito-eindtoets vooral ook aan de bovenkant van die verdeling moet onderscheiden. Stijging van scores op de PRIMA-taaltoets wijzen mogelijk vooral op vooruitgang bij de minder presterende leerlingen.

Voorzover in de eerdere discussie (zie de eerste paragraaf van dit artikel) wel werd gezegd dat een (vermeende) daling van de Cito-scores te maken had met een (vermeende) afnemende kwaliteit van het basisonderwijs en/of de leerkrachten in Nederland, vinden we in dit onderzoek dus wat anders: de daling van de Cito-scores is slechts gering en niet doorgaand; en andere indicatoren van de prestaties en capaciteiten van leerlingen geven een ander beeld.

We zouden willen aanbevelen om uitspraken over de kwaliteit van het basisonderwijs in Nederland niet louter te baseren op trends over slechts enkele meetmomenten in gemiddelde scores op de eindtoets basisonderwijs van het Cito, maar om daar meerdere indicatoren gemeten over een groter aantal jaren voor te gebruiken.

Met het bestaan van het Regulier en Integraal Schooltoezicht van de Inspectie en van onderzoeken als het PRIMA-cohortonderzoek zijn ook meerdere van zulke indicatoren in principe beschikbaar. En met een toename van het gebruik van leerlingvolgsystemen (Roeleveld, Otter & Blok, 2001; Blok, Otter & Roeleveld, 2001) wordt ook op scholen zelf een schat aan gegevens verzameld, die mogelijk in de toekomst een gevarieerder en rijker beeld van de ontwikkeling van het basisonderwijs in Nederland mogelijk kunnen maken.

## Noten

- 1 "Trouw", 2 februari 2001, "Eindniveau school daalt".
- 2 "Trouw", 6 februari 2001, "Cito-scores verslechterd".
- 3 De gegevens beslaan daarmee een langere tijdsspanne dan in het recente artikel van Webink (2002).
- 4 Onze selectie van gegevens is daarmee anders dan die van Webink (2002). Deze maakt ook gebruik van de aanvullende steekproef van PRIMA en heeft daarom een oververtegenwoordiging van scholen met relatief veel kansarme leerlingen. Dit beperkt de generaliseerbaarheid van zijn bevindingen.
- 5 Bij de scores op de verschillende onderdelen van de Cito-eindtoets wordt een dergelijke procedure niet gehanteerd, zodat daarbij een vergelijking van gemiddelde scores over de jaren niet zinvol is. Bij de eindtoets van 2001 is een ander zgn. "moederjaar" gebruikt dan bij de eerdere jaren. D.m.v. een omrekenformule kunnen de standaardcores uit de eerdere jaren weer vergelijkbaar gemaakt worden met de scores uit 2001.
- 6 Preciezer geformuleerd: zij hebben gegevens over de uitslag van de eindtoets opgeleverd aan de PRIMA-onderzoekers. Deze gegevens worden pas na afloop van het eigenlijke PRIMA-veldwerk bij de scholen opgevraagd. De respons is over de verschillende jaren in de orde van 80%.
- 7 De enige significante interactieterm in Tabel 4 voor leerlingen van hoger opgeleide ouders in 2001 blijkt in dit restrictievere model niet meer significant en kan verder worden weggelaten.
- 8 De variabelen *regio* en *stad* blijven over de jaren gelijk en zijn daarmee kenmerken op het niveau van de school. De andere variabelen kunnen van jaar tot jaar wisselen en zijn daarmee in het multivolumemodel kenmerken van het niveau jaar.
- 9 Deze percentages zijn berekend over het aantal PRIMA-leerlingen in de groepen 2, 4, 6 en 8 waarvan de sociaal-etnische herkomst bekend is; leerlingen met onbekende herkomst zijn niet meegerekend.
- 10 Zie "Trouw", 2 februari 2001, "Eindniveau school daalt".
- 11 Noord-Oost = Groningen, Friesland, Drenthe en Overijssel; Midden-West = Flevoland, Utrecht, Noord en Zuid Holland; Zuid = Zeeland, Noord Brabant, Gelderland en Limburg.
- 12 De Cito-eindtoets wordt doorgaans in februari af-

genomen; de PRIMA-toetsingen vinden plaats tussen begin januari en eind maart.

- 13 Bij Cohen (1988) worden effecten van 0.20 'weak' genoemd, effecten van 0.50 'medium' en effecten van 0.80 'strong'.

## Literatuur

- Blok, H., Otter M.E., & Roeleveld, J. (2001). *Het gebruik van leerlingvolgsystemen anno 2000*. SCO-rapport 626. Amsterdam: SCO-Kohnstamm Instituut.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- 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. Langen, A. van, & Oudenhoven, D. (1994). *De toetsen voor het cohort Primair Onderwijs*. Nijmegen, ITS
- Driessen, G. Langen, A. van, & Vierke, H. (2002). *Basisonderwijs: veldwerkverslag, leerlinggegevens en oudervragenlijst. Basisrapportage PRIMA-cohortonderzoek. Vierde meting 2000/2001*. Nijmegen, ITS
- Dronkers, J. (1990) De ontwikkelingen in het schoolloopbaanonderzoek: een terugblik op een decennium. *Tijdschrift voor Onderwijsresearch*, 15, 8-22
- Dronkers, J., & Ultee, W.C. (Eds.) (1995) *Verschuivende ongelijkheid in Nederland: Sociale laagheid en mobiliteit*. Assen: Van Gorcum
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold
- Jungbluth, P., Langen, A. van, Peetsma, Th., & Vierke, H. (1996). *Leerlinggegevens basisonderwijs en speciaal onderwijs. Technische rapportage PRIMA-cohortonderzoek 1994/95*. Nijmegen/Amsterdam: ITS/SCO-Kohnstamm Instituut
- Kamphuis, F., Mulder, L., Vierke, H., Overmaat, M., & Koopman, P. (1998). *De relatie tussen PRIMA-toetsen en toetsen uit het CITO-leerlingvolgsysteem*. Arnhem/Nijmegen/Amsterdam: CITO, ITS, SCO-Kohnstamm Instituut
- Langen, A. van, Viere, H., & Robijns, M. (1996). *Veldwerkverslag basisonderwijs en speciaal onderwijs. Technische rapportage PRIMA-cohortonderzoek 1994/95*. Nijmegen/Amsterdam: ITS/SCO-Kohnstamm Instituut
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D.,

Langford, I., & Lewis, T. (2000) *A user's guide to MlwiN*. London: Multilevel Models Project, University of London

Roeleveld, J. (2001). *De Cito eindtoets basisonderwijs binnen PRIMA van 1995 tot en met 2001*. Amsterdam: SCO-Kohnstamm Instituut

Roeleveld, J., Otter M.E., & Blok, H. (2001). *Leerlingvolgsystemen in de jaren negentig. Secundaire analyses op gegevens uit PRIMA en IST*. SCO-rapport 625. Amsterdam: SCO-Kohnstamm Instituut

Roeleveld, J., & Meijden, A. van der (2002). *Speciaal basisonderwijs: veldwerkverslag, leerlinggegevens en oudervragenlijst. Basisrapportage PRIMA-cohortonderzoek. Vierde meting 2000/2001*. Amsterdam: SCO-Kohnstamm Instituut

Snijders, T.A.B., & Bosker, R.J. (1999). *Multilevel analysis. An Introduction to basic and advanced multilevel modeling*. Newbury Park: Sage.

Uiterwijk, J.H., & Theunissen, T.J.J.M. (2001) *Verantwoording eindtoets basisonderwijs 1997*. Arnhem: Citogroep.

Vierke, H., & Mulder, L. (2001) Daling van de Cito-scores? Vergelijkbaarheid van de cijfers over de jaren heen. Bijlage bij: Roeleveld, J. (2001). *De Cito-eindtoets basisonderwijs binnen PRIMA van 1995 tot en met 2001*. Amsterdam: SCO-Kohnstamm Instituut

Webbink, D. (2002) Moeten we ons zorgen maken over dalende scores op de Eindtoets Basisonderwijs? *Pedagogische Studiën*, 79, 184-191.

Manuscript aanvaard: 9 september 2002

## Auteur

**Jaap Roeleveld** is onderzoeker bij het SCO-Kohnstamm Instituut van de Faculteit der Maatschappij- en Gedragswetenschappen, Universiteit van Amsterdam.

*Correspondentieadres:* Jaap Roeleveld, Universiteit van Amsterdam, SCO-Kohnstamm Instituut, Postbus 94208, 1090 GE Amsterdam, e-mail: jaapr@educ.uva.nl

## Abstract

### **Quality of primary education: are Cito scores going down?**

Though educational levels of parents are rising during the last years, average scores on a general performance test at the end of primary education (*Cito test*) remain unchanged, thus indicating a possible descent of quality in Dutch primary education. This study investigates whether average scores are decreasing over the years 1995 to 2001, when controlling for the educational levels of parents. Data from the PRIMA cohort study are analyzed using three level models in which children are nested within years within schools.

The results indicate a minor tendency of decline of average Cito scores over the years and it is shown that this tendency cannot be attributed to a change in schools taking the Cito test. Scores on other performance and ability tests, however, show a tendency to rise over the same years. So it is argued that the quality of primary education should not be evaluated on one output measure alone, but that more aspects of education should be taken into account.