

Moeten we ons zorgen maken over dalende scores op de Eindtoets Basisonderwijs?

D. Webbink¹

Samenvatting

In dit artikel wordt de ontwikkeling van de scores op de Eindtoets Basisonderwijs van het Cito geanalyseerd tussen 1995 en 1999. Daarvoor wordt gebruikt gemaakt van gegevens van leerlingen uit het PRIMA-onderzoek. De steekproef is afgebakend tot scholen die in de onderzochte jaren steeds deelnamen aan het PRIMA-onderzoek en aan de Cito-eindtoets. Voorts is rekening gehouden met zoveel mogelijk beschikbare kenmerken van leerlingen en het feit dat prestaties van leerlingen op een bepaalde school kunnen samenhangen. Het belangrijkste resultaat van de analyses is dat het prestatieniveau op de Cito-eindtoets in de tijd niet veranderd is, noch dat er verschillen in ontwikkeling zijn tussen “witte” en “zwarte” leerlingen.

1 Inleiding

Begin vorig jaar was er in de media enig rumoer over een mogelijke daling van de scores op de Eindtoets Basisonderwijs van het Cito.² Het onderwerp van discussie was of de Citoscores al of niet dalen en voor welke groepen dit het geval was.³ Voor de gehele leerlingenpopulatie is, volgens gegevens van het Cito, de daling niet groot: de toetsresultaten dalen met ongeveer een half procent.⁴ Als rekening wordt gehouden met kenmerken van individuele leerlingen, gemeten in het PRIMA-onderzoek, zou echter een ander beeld ontstaan: “witte” leerlingen scoren slechter, Turkse en Marokkaanse leerlingen gaan beter presteren. Een daling van de Citoscores kan een aanwijzing zijn voor een verslechterende kwaliteit van het Nederlands basisonderwijs en is derhalve reden tot bezorgdheid.⁵ Het is echter ook mogelijk dat de daling samenhangt met andere factoren. Zo kunnen verschillen in de steekproef van scholen en leerlingen of veranderingen in buitenschoolse factoren de resultaten op de

Citotoets beïnvloeden.⁶ Ook de ijking van de Citotoetsen zou een rol kunnen spelen. Dit laatste lijkt echter niet waarschijnlijk, omdat hiervoor een zorgvuldig ontworpen equivalentieprocedure wordt gebruikt.⁷

In dit artikel wordt een nieuwe analyse uitgevoerd naar de ontwikkeling van de Cito-scores tussen 1995 en 1999. Overeenkomstig de eerdere analyses die aanleiding gaven voor de discussie, wordt gebruikt gemaakt van de PRIMA-data. Het doel van de analyse is om vast te stellen of de Citoscores daadwerkelijk dalen en of er verschillen zijn tussen groepen als wordt gecontroleerd voor zoveel mogelijk factoren die niet gerelateerd zijn aan de kwaliteit van het onderwijs. In de analyse worden allereerst de scholen constant gehouden, dat wil zeggen: de steekproef van scholen die deelnemen aan de Eindtoets in jaar t is gelijk aan die in jaar $t + 2$.⁸ Daarnaast wordt rekening gehouden met zoveel mogelijk relevante kenmerken van leerlingen en het feit dat prestaties van leerlingen op een bepaalde school kunnen samenhangen.

2 Aanpak van de analyse

De meest directe aanpak om de verschillen tussen de jaren te bepalen is het schatten van een regressiemodel met de *Citoscore* als afhankelijke variabele en als onafhankelijke variabelen een reeks relevante kenmerken van de leerlingen en het jaar waarin de toets is afgenomen (Vergelijking 1):

$$Y_{ijt} = \alpha + X_i \beta + T\delta + v_{ijt}$$

waarbij:

Y_{ijt}	Citoscore van leerling i op school j in jaar t .
T	jaardummy
X_i'	kenmerken van leerlingen (als getransponeerde kolomvector)

v_{ijt} stoorterm van niet waargenomen factoren met verwachte waarde nul en variantie σ^2
 α, β, δ te schatten coëfficiënten

In dit model wordt bij het vaststellen van het verschil in Citoscores tussen de jaren rekening gehouden met verschillen in de samenstelling van de leerlingenpopulatie. Leerlingen in jaar t kunnen bijvoorbeeld gemiddeld hogere scores dan leerlingen in jaar $t + 2$ als zij een hogere sociaal-economische achtergrond hebben. In Model 1 wordt echter geen rekening gehouden met de clustering van leerlingen op een bepaalde school in een bepaald jaar. De prestaties van leerlingen op een bepaalde school kunnen beïnvloed worden door schoolspecifieke factoren die niet worden waargenomen. In dat geval is sprake van een samengestelde storingsterm met een schoolspecifiek en een leerspecifiek deel (Vergelijking 2):

$$Y_{ijt} = \alpha + X_i \beta + T\delta + e_{jt} + u_{ijt}$$

met $v_{ijt} = e_{jt} + u_{ijt}$, e_{jt} zijn niet waargenomen schoolspecifieke factoren en u_{ijt} zijn individuspecifieke niet waargenomen factoren. Als de storingstermen van leerlingen op dezelfde school in hetzelfde jaar gecorreleerd zijn ($Corr(v_{bjt}, v_{cjt}) = \sigma_e^2 / (\sigma_e^2 + \sigma_u^2) \geq 0$ met $\sigma_e^2 = Var(e_{jt})$ en $\sigma_u^2 = Var(u_{ijt})$) worden de standaardfouten van de coëfficiënten in Vergelijking 1 niet correct geschat. Met een 'random-effects'-model kan hiermee rekening worden gehouden.

Vergelijking 2 wordt een 'random-effects'-model als wordt aangenomen dat de schoolspecifieke storingsterm (e_{jt}) niet samenhangt met de onafhankelijke variabelen ($Cov(X_i, e_{jt}) = 0$).⁹ Door het model te schatten met een gegeneraliseerde kleinste-kwadratentechniek wordt de schoolspecifieke factor (e_{jt}) geëlimineerd.¹⁰ De standaardfouten van de coëfficiënten worden in dat geval wel correct geschat. Deze nemen toe als de storingstermen van leerlingen op dezelfde school gecorreleerd zijn.

In het model wordt verder geen rekening gehouden met verschillen tussen scholen. Enerzijds is dat ook niet de bedoeling, want de mogelijke afname van de onderwijskwaliteit wordt gemeten met de jaardummy. Anderzijds kunnen steekproefverschillen de resultaten vertekenen. De hogere score van leerlingen in jaar t kan bijvoorbeeld worden veroorzaakt doordat de steekproef in dat jaar relatief meer leerlingen bevat van kwalitatief goede scholen dan de steekproef in jaar $t + 2$. Ook is het mogelijk dat leerlingen beter of slechter presteren, omdat hun school al of niet eerdere ervaring heeft opgedaan met de Citotoets. Scholen die al vaker aan de toets hebben meegedaan kunnen hun leerlingen daar wellicht beter op voorbereiden.¹¹ Het is daarom van belang om in beide jaren een vergelijkbare steekproef van scholen te hebben. Bij de analyse is ervoor gekozen om de steekproef af te bakenen tot scholen die in de jaren die worden vergeleken steeds hebben deelgenomen aan de Citotoets (en aan het PRIMA-onderzoek). Deze afbakening heeft als bijkomend voordeel dat daarmee waarschijnlijk ook het verstorende effect wordt geëlimineerd van niet waargenomen kenmerken die verband houden met de keuze van een school. De leerlingenpopulatie op een school komt namelijk niet door toeval tot stand. Ouders kiezen een basisschool bijvoorbeeld vanwege de afstand, de pedagogisch-didactische aanpak of de samenstelling van de leerlingenpopulatie. Meer gemotiveerde ouders hebben wellicht een voorkeur voor een bepaalde school. Dit type kenmerken wordt niet waargenomen door de onderzoeker; ze kunnen evenwel wel leiden tot verschillen in prestaties op de Citotoets. Door steeds dezelfde scholen te nemen voor de jaren die worden vergeleken, wordt rekening gehouden met deze niet waargenomen selectiviteit.¹² Een nadeel van deze afbakening is dat de representativiteit van de resultaten kan verminderen. Het gaat immers om een selectie van scholen uit een representatieve steekproef. Deze selectie vergroot echter de validiteit van de resultaten. In feite is sprake van een 'trade-off' tussen (interne) validiteit en representativiteit (externe validiteit). De (interne) validiteit weegt in de navolgende analyses het zwaarst, hetgeen

betekent dat niet zeker is of de gevonden resultaten geldig zijn voor het gehele Nederlandse basisonderwijs. Overigens is het de vraag wat representativiteit waard is als het streven ernaar de validiteit van de resultaten ondermijnt.

3 Data en selectie van variabelen

In de analyse wordt gebruik gemaakt van de gegevens uit het PRIMA-onderzoek. Dit is een longitudinaal onderzoek onder leerlingen uit het basisonderwijs in groep 2, 4, 6 en 8. In het onderzoek wordt informatie verzameld over leerlingen, docenten, directies, scholen en ouders.¹³ We maken gebruik van drie golven van het PRIMA-onderzoek.¹⁴ Dit betreft de schooljaren 1994-1995, 1996-1997 en 1998-1999. De Cito-eindtoets wordt afgenomen bij leerlingen uit groep 8, meestal in januari of februari. In het vervolg zullen we de schooljaren steeds aanduiden met het jaar van de afname van de Cito-eindtoets.

De afhankelijke variabele in de analyse is de *Cito-eindtoets*, dit is de zogenoemde eindtoets van het basisonderwijs. De totaalscore op deze toets varieert van 501 tot en met 550. Deze totaalscore is samengesteld uit drie onderdelen: taal, rekenen en informatieverwerking. Bij elk onderdeel variëren de scores van 0 tot en met 60.¹⁵

Als onafhankelijke variabelen zijn indicatoren van de sociaal-economische achtergrond opgenomen: *SES*, *gezinsamenstelling* en *verblijfsduur in Nederland*. De *SES*-variabele is geconstrueerd in het PRIMA-onderzoek als combinatie van het opleidingsniveau van de ouders en de etniciteit. In de analyse worden de categorieën als dummy gebruikt:

- *lbo/Tur/Mar*: beide ouders hebben maximaal een lbo-opleiding en zijn van Turkse of Marokkaanse herkomst;
- *lbo/ov*: zelfde, herkomst anders dan Turks, Marokkaans of Nederlands;
- *lbo/Ned*: zelfde, herkomst Nederlands;
- *mbo*: de hoogst opgeleide ouder heeft maximaal een mbo-opleiding;
- *hbo/wo*: de hoogst opgeleide ouder heeft een hbo- of w.o.-niveau;
- *SES onbekend*: de SES-score is onbekend. De samenstelling van het gezin is weergege-

ven in vier categorieën: *vader en moeder aanwezig*, *alleen moeder*, *alleen vader*, *anders/onbekend*. Deze variabele is niet beschikbaar in 1995. De verblijfsduur in Nederland is gemeten in vijf categorieën: *minder dan 1 jaar*, *1-3 jaar*, *4-5 jaar*, *meer dan 5 jaar*, *altijd*. Ook deze variabele is niet beschikbaar in 1995. Daarnaast zijn het *geslacht* en de *leeftijd* van de leerling opgenomen. Het PRIMA-onderzoek bevat, als gevolg van het panelkarakter, ook informatie over eerdere schoolvorderingen, zoals scores op twee jaar eerder afgenomen taal- en rekenoetsen. Deze indicatoren van eerdere schoolvorderingen zijn echter voor de onderhavige analyses niet goed bruikbaar. De taal- en rekenoetsen zijn in de eerste golf van het PRIMA-onderzoek namelijk afgenomen in de periode oktober tot en met december. In de tweede golf van het onderzoek zijn de toetsen afgenomen tussen januari en maart. Dit betekent dat leerlingen van de tweede golf gemiddeld drie maanden langer op school hebben gezeten, waardoor hun scores niet goed vergelijkbaar zijn met die van leerlingen uit de eerste golf. In het PRIMA-onderzoek wordt ook een intelligentietest afgenomen, bestaande uit de non-verbale onderdelen Figuren Samenstellen en Exclusie. Deze variabele is niet gebruikt in de analyse, omdat niet duidelijk is in hoeverre de scores op deze test onafhankelijk zijn van het gevolgde onderwijs.¹⁶ In Tabel 1 zijn enkele kerngegevens weergegeven van de variabelen voor de deelsteekproef die in de analyse is gebruikt (scholen die in beide jaren aan PRIMA hebben meegedaan en aan de Eindtoets Basisonderwijs van het Cito).

Een eerste blik op Tabel 1 leert dat de gemiddelde score op de Eindtoets Basisonderwijs niet daalt als de steekproef wordt afgebakend tot scholen die in de jaren die worden vergeleken steeds hebben meegedaan aan de toets. De gemiddelde scores liggen lager dan de populatiegemiddelden (noot 4), omdat in de gekozen steekproef leerlingen uit de doelgroepen van het achterstandenbeleid oververtegenwoordigd zijn.

Tabel 1

Beschrijving van geselecteerde variabelen in deelsteekproef van scholen die deelnemen aan de Cito-toets in vergeleken jaren

	1995		1997		1999	
	Gemid.	St. dev.	Gemid.	Std. dev.	Gemid.	Std. dev.
Cito eindtoets	532.8	10.6	532.9	10.3	533.0	10.1
Leeftijd	12.0	0.5	12.0	0.5	12.0	0.5
Periode in Ned (1-5)	Nb		4.9	0.6	4.9	0.7
		%		%		%
Meisje		52.0		49.7		50.9
SES ouders						
Lbo/ Tur/Mar		9.5		11.9		11.0
Lbo/ overig		3.7		5.2		5.4
Lbo Ned		29.5		32.4		28.9
Mbo		26.0		30.0		31.8
Hbo/Wo		15.3		15.3		18.1
SES onbekend		16.0		5.2		4.7
2 ouders in gezin		Nb		82.9		84.1
Alleen moeder		Nb		9.4		8.9
Alleen vader		Nb		0.8		0.6
Aantal leerlingen		4856		4803		5110

Noot. Nb= niet beschikbaar

Tabel 2

'Random'-effectschatting van scores op Cito-eindtoets basisonderwijs '95-'99, gepoolde steekproeven van twee jaar

	1995&1999		1995&1997		1997&1999	
	Coeffi.	Std. fout	Coeffi.	Std. fout	Coeffi.	Std. fout
Jaar t.o.v. 1995	-0.04	0.49 n.s.	-0.32	0.42 n.s.		
Jaar t.o.v. 1997					-0.22	0.40 n.s.
Meisje	-0.66	0.22 **	-1.00	0.19 **	-0.14	0.19 n.s.
Leeftijd	-4.65	0.23 **	-4.53	0.20 **	-4.36	0.20 **
SES (ref.=Mbo)						
Lbo/Tur/Mar	-6.16	0.49 **	-6.94	0.42 **	-6.40	0.40 **
Lbo/overig	-3.86	0.63 **	-4.39	0.52 **	-3.68	0.49 **
Lbo/Ned	-3.90	0.30 **	-4.21	0.25 **	-4.01	0.24**
Hbo/wo	3.29	0.35 **	2.80	0.29 **	3.59	0.28 **
Ses onbekend	-4.42	0.51 **	-4.33	0.42 **	-3.31	0.63 **
Constante	591.25	2.82 **	590.70	2.39 **	586.91	2.43 **
R ² -totaal	0.19		0.19		0.20	
R ² -binnen	0.15		0.15		0.15	
R ² -tussen	0.35		0.31		0.37	
Aantal leerlingen	6271		8941		8506	
Aantal scholen	170		257		239	

Noot. ** sign. op 1 %-niveau, * sign. op 5 %-niveau, + significant op 10 %-niveau, n.s. niet significant.
 'Random'-effecten zijn opgenomen per school per jaar. R²-totaal is de bekende R-kwadraat, R²- binnen geeft de verklaarde variantie binnen scholen per jaar en R²- tussen geeft de verklaarde variantie tussen scholen per jaar.

4 Resultaten

Op grond van de beschikbaarheid van de gegevens is de analyse uitgevoerd in twee stappen. Allereerst zijn vergelijkbare modellen geschat voor de gehele periode 1995-1999 en voor de twee deelperiodes. Vervolgens zijn voor de tweede deelperiode (1997-1999) meer uitgebreide modellen geschat voor de eindtoets. Voor deze periode zijn namelijk meer onafhankelijke variabelen beschikbaar.

In Tabel 2 zijn de resultaten weergegeven van het 'randomeffects-model' voor drie periodes: 1995-1999, 1995-1997 en 1997-1999. Voor alle jaren zijn dezelfde onafhankelijke variabelen opgenomen. In de kolom 1995-1999 wordt de analyse uitgevoerd op een steekproef van scholen die zowel in 1995 als in 1999 deelnamen aan de Cito-toets. Hetzelfde geldt voor de twee laatste kolommen. Het gaat derhalve steeds om gepoolde steekproeven van twee jaar. Daarmee wordt een zo groot mogelijk aantal scholen vergeleken.

De eerste twee rijen met schattingsresultaten laten zien dat in geen van de drie periodes sprake is van een significante daling van de scores op de Eindtoets. De geschatte coëfficiënten voor de andere variabelen zijn vrij stabiel. Leerlingen met een hogere *SES* scoren hoger, oudere leerlingen scoren lager. Meisjes scoren alleen in de eerste periode lager dan jongens. In de discussie, die in de inleiding is genoemd, ging het ook om de vraag of er verschillen in ontwikkeling waren voor "witte" en "zwarte" leerlingen. Om na te gaan of de verandering in de toetsscores afhangt van de sociaal-economische achter-

grond van de leerlingen, zijn ook modellen geschat waarin de interactie tussen sociaal-economische achtergrond (*SES*) en periode (*T*) is opgenomen (Vergelijking 3)¹⁷:

$$Y_{ijt} = \alpha + X_{ij}\beta + T\delta + T.SES\gamma + e_{jt} + u_{ijt}$$

In Tabel 3 zijn de schattingsresultaten voor deze interacties gegeven. De resultaten laten zien dat de interacties tussen tijd en *SES*-categorie nauwelijks significante resultaten opleveren. Dit betekent dat de verandering van de scores op de Cito-eindtoets voor een bepaalde *SES*-categorie niet afwijkt van de algehele tijdstrend die in de eerste twee rijen is aangegeven. Alleen leerlingen met een onbekende *SES* gaan beter presteren. Het is echter onduidelijk wat dit betekent; we weten bijvoorbeeld niet of de samenstelling van deze categorie leerlingen is veranderd en bijvoorbeeld een hogere *SES* heeft gekregen.

Verschillen tussen 1997 en 1999

Voor de analyse van de ontwikkeling tussen 1997 en 1999 is ook informatie beschikbaar over de samenstelling van het gezin en de verblijfsduur in Nederland. Tabel 4 geeft de resultaten van het randomeffects-model waarin deze variabelen worden gebruikt. Daarnaast is een tweede model geschat met interacties van *SES* en *Tijd* (*T*).

Ook in de modellen met een uitgebreidere set van achtergrondkenmerken wordt geen significante daling op de Cito-eindtoets gevonden. De geschatte coëfficiënten voor de interacties duiden er bovendien niet op dat de ontwikkeling in de Cito-score afhangt van de

Tabel 3

Modellen als in Tabel 2 aangevuld met interacties van SES en tijd

	1995&1999		1995&1997		1997&1999	
	Coef. f.	Std. fout	Coef. f.	Std. fout	Coef. f.	Std. fout
Jaar t.o.v. 1995	-0.49	0.61 n.s.	-0.22	0.52 n.s.		
Jaar t.o.v. 1997					-0.80	0.49 n.s.
SES (ref.=Mbo)						
Lbo/Tur/Mar*tijd	0.36	0.98 n.s.	-0.45	0.83 n.s.	1.09	0.79 n.s.
Lbo/ov*tijd	1.95	1.28 n.s.	0.56	1.04 n.s.	1.00	0.97 n.s.
Lbo/Ned*tijd	0.17	0.60 n.s.	-0.25	0.49 n.s.	0.56	0.49 n.s.
Hbo/wo*tijd	0.82	0.70 n.s.	-0.31	0.59 n.s.	0.60	0.57 n.s.
Ses/onb*tijd	2.23	1.17 +	0.74	0.95 n.s.	3.05	1.25 *

Noot. ** sign. op 1 %-niveau, * sign. op 5 %-niveau, + significant op 10 %-niveau, n.s. niet significant.

Tabel 4

'Random'-effectschatting van verandering op de Cito-eindtoets '97/'99, uitgebreide modellen

	Coeffi.	Std. fout	Coeffi.	Std. fout
'99 t.o.v. '97	-0.11	0.40 n.s.	-0.69	0.49 n.s.
Meisje	-0.04	0.19 n.s.	-0.04	0.19 n.s.
Leeftijd	-4.34	0.21 **	-4.34	0.21 **
SES (ref=Mbo)				
Lbo/Tur/Mar	-6.29	0.41 **	-6.90	0.60 **
Lbo/overig	-3.52	0.52 **	-4.06	0.73 **
Lbo/Ned	-3.97	0.25 **	-4.24	0.36 **
Hbo/w.o.	3.62	0.29 **	3.29	0.43 **
Ses/onb	-3.29	0.70 **	-5.49	1.01 **
Gezin (ref= moeder +vader)				
Alleen moeder	-0.12	0.35 n.s.	-0.11	0.35 n.s.
Alleen vader	-0.38	1.18 n.s.	-0.49	1.18 n.s.
Verblijfsduur (ref = altijd)				
< 1 jaar	-1.95	1.14 +	-1.93	1.14 +
1-3 jaar	-1.88	0.84 +	-1.98	0.84 +
4-5 jaar	0.55	1.03 n.s.	0.57	1.03 n.s.
> 5 jaar	-0.93	0.66 n.s.	-0.95	0.66 n.s.
Interactie SES en tijd (ref.=Mbo)				
Lbo/Tur/Mar*tijd			1.09	0.81 n.s.
Lbo/ov*tijd			1.03	0.99 n.s.
Lbo/Ned*tijd			0.50	0.50 n.s.
Hbo/w.o.*tijd			0.61	0.57 n.s.
Ses/onb*tijd			4.22	1.41 **
Constante	586.49	2.55 **	586.76	2.55 **
R ² -totaal	0.20		0.20	
R ² - binnen	0.15		0.15	
R ² - tussen	0.38		0.39	
Aantal leerlingen	8061		8061	
Aantal scholen	234		234	

Noot. ** sign. op 1 %-niveau, * sign. op 5 %-niveau, + significant op 10 %-niveau, n.s. niet significant.

sociaal-economische achtergrond van de leerlingen. De verblijfsduur in Nederland lijkt samen te hangen met de score op de Cito-eindtoets. Leerlingen die minder dan drie jaar in Nederland verblijven, scoren lager op de Cito-eindtoets. De score van leerlingen die vier jaar of langer in Nederland wonen, wijkt niet significant af van die van leerlingen die altijd in Nederland hebben gewoond. Uiteraard is het de vraag in hoeverre deze effecten te maken hebben met de duur van het verblijf, of dat andere factoren - die zowel de verblijfsduur als de scores op de Cito-toets beïnvloeden - een rol spelen. De onderhavige analyse kan deze vraag niet be-

antwoorden, omdat de variabele *verblijfsduur* in Tabel 4 slechts als controlevariabele is opgenomen.¹⁸ Tot slot worden geen verschillen gevonden tussen leerlingen uit “complete gezinnen” en leerlingen met alleen een moeder of vader in het gezin.

5 Conclusie

In dit onderzoek wordt geen daling gevonden van de gemiddelde score op de Eindtoets Basisonderwijs, noch tussen 1995 en 1997, noch tussen 1997 en 1999. Evenmin wordt gevonden dat “witte” leerlingen zwakker zijn

gaan presteren en dat Turkse en Marokkaanse leerlingen beter zijn gaan presteren. Anders dan in eerder onderzoek naar de recente ontwikkeling van de Citoscore, is de steekproef in het onderhavige onderzoek afgebakend tot scholen die in de onderzochte jaren steeds deelnamen aan het PRIMA-onderzoek en aan de Cito-eindtoets. Deze afbakening versterkt de interne validiteit van het onderzoek omdat versturende effecten door verschillen in de steekproef van scholen worden geëlimineerd. Nadeel is dat dit mogelijk ten koste gaat van de representativiteit van de bevindingen, het gaat immers om een selectie van scholen uit een representatieve steekproef, en dat dit de geldigheid van de bevindingen beperkt. Deze beperking lijkt echter niet groot omdat het hier gevonden beeld niet afwijkt van de ontwikkeling van het gemiddelde voor de gehele populatie, zoals gemeten door het Cito. De bevindingen uit dit onderzoek geven, anders dan de berichten in de media van vorig jaar, geen aanleiding tot bezorgdheid over de kwaliteit van het Nederlandse basisonderwijs.

Noten

- 1 De auteur dankt Hessel Oosterbeek, Jaap Roeleveld en beide reviewers voor commentaar op een eerdere versie en Henny Uiterwijk voor informatie over de Cito-toets.
- 2 De directe aanleiding was een artikel in dagblad "Trouw", gebaseerd op resultaten van onderzoek door dr. P. Jungbluth. Op de Onderwijs Research Dagen 2001 werden nieuwe resultaten gepresenteerd die de daling bevestigen, met name bij kinderen van middelbaar en hoger opgeleide ouders (Roeleveld, 2001). Diverse verklaringen werden aangedragen, zoals de afnemende kwaliteit van Pabo-opleidingen of een meer bewuste en beter geïnformeerde schoolkeuze door lager opgeleide ouders.
- 3 Met Citoscore wordt steeds bedoeld de score op de Eindtoets Basisonderwijs.
- 4 In 1995 was het populatiegemiddelde 534.88, in 1997 was dit 534.93 en in 1999 was dit 534.60.
- 5 De Eindtoets Basisonderwijs van het Cito heeft twee functies: in de eerste plaats verschaft de toets informatie over individuele leerlingen ten behoeve van de beslissing over het vervolg-

onderwijs. In de tweede plaats levert de toets "informatie ten behoeve van de evaluatie van het onderwijs" (Uiterwijk & Theunissen, 2001).

- 6 Voor het belang van buitenschoolse factoren en hun interactie met schoolfactoren, zie Todd en Wolpin (2002).
- 7 De kern van deze procedure is dat elk jaar een steekproef van leerlingen dezelfde toetsen maakt (ankertoetsen). Met behulp van de scores op deze toetsen worden de resultaten van de Eindtoets Basisonderwijs genormeerd (Uiterwijk & Theunissen, 2001).
- 8 De PRIMA-data worden tweejaarlijks verzameld.
- 9 In de onderwijskundige literatuur wordt dit model ook wel aangeduid als 'multilevel model' (zie bijv. Snijders & Bosker, 1999). In de econometrie wordt dit type modellen gerekend tot de panel data modellen. Zonder de aanname van onafhankelijkheid van de schoolspecifieke storings-term en de overige onafhankelijke variabelen wordt gesproken van een 'fixed effects model'.
- 10 Zie bijv. Wooldridge (2000), pp. 450-451.
- 11 Volgens staatssecretaris Adelmund oefenen veel scholen op het maken van de Cito-toets (toespraak op conferentie "Het oog der natie", Amsterdam 26 september 2001).
- 12 De vertekening door niet-waargenomen selectiviteit wordt in de onderwijseconomie als een belangrijk probleem beschouwd bij het schatten van causale effecten. Voor een overzicht van technieken om hiermee rekening te houden, zie Angrist en Krueger (1999).
- 13 Het PRIMA-onderzoek wordt uitgevoerd door het ITS te Nijmegen en het SCO-Kohnstamm Instituut te Amsterdam. Over opzet en resultaten van het onderzoek zijn diverse technische en basis-rapportages beschikbaar.
- 14 Inmiddels is ook de vierde golf uitgevoerd. Deze data zijn evenwel nog niet beschikbaar voor analyse.
- 15 De ontwikkeling van de scores op de deelttoetsen wordt niet geanalyseerd, omdat deze toetsen niet geïjkt worden.
- 16 Onderzoek in de Verenigde Staten laat bijvoorbeeld zien dat een exogene toename van het gevolgde onderwijs leidt tot een hogere IQ-score (Neal & Johnson, 1996).
- 17 SES maakt ook deel uit van X_i' .
- 18 Een beschrijving van methoden voor het vaststellen van causale effecten is te vinden in Angrist en Krueger (1999).

Literatuur

- Angrist, J.D., & Krueger, A.B. (1999). Empirical strategies in labor economics. In O. Ashenfelter, & D. Card (Eds.), *Handbook of Labor Economics* (pp. 1277-1366). Amsterdam: North Holland.
- Eindniveau school daalt. (2001, 2 februari). *Trouw*.
- Neal, D.A., & Johnson, W.R. (1996). The role of pre-market forces in black-white wage differences. *Journal of Political Economy*, 104(5), 869-895.
- Roeleveld, J. (2001), *Multilevel blik naar historische ontwikkeling: de Cito eindtoets basisonderwijs binnen Prima*. Concept-paper. Amsterdam: SCO-Kohnstamm Instituut.
- Snijders, T.A.B., & Bosker, R.J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modelling*. London: Sage.
- Todd, P.E., & Wolpin, K.I. (2002). Towards a unified approach for modelling the production function for cognitive achievement. *Economic Journal*, in press.
- Uiterwijk, J.H., & Theunissen, T.J.J.M. (2001). *Verantwoording eindtoets basisonderwijs 1997*. Arnhem: Cito.
- Wooldridge, J.M. (2000). *Introductory econometrics, a modern approach*. Mason, USA: South-Western College Publishing.

Manuscript aanvaard: 27 maart 2002

Auteur

Dinand Webbink is als senior onderzoeker verbonden aan het NWO-prioriteitsprogramma SCHOLAR dat aan de Universiteit van Amsterdam wordt uitgevoerd. Daarnaast is hij werkzaam bij de afdeling Kenniseconomie van het CPB.

Correspondentieadres: D. Webbink, SCHOLAR,
Roetersstraat 11, 1018 WB Amsterdam,
e-mail: webbink@cpb.nl

Abstract

Do we have to worry about declining test scores at the end of primary education?

Several articles in daily newspapers suggest declining scores on a nation wide voluntary test for pupils at the end of primary education. Moreover, performance of white pupils deviates negatively from performance of ethnic minorities. In this article data from three waves of a large longitudinal research project in primary education for the period 1995-1999 have been analysed. In contrast to previous analyses the data have been restricted to schools which participated both in the longitudinal project and in the nation wide test in each of the years that are compared. No evidence has been found for declining test scores nor for a weakening of the performance of white pupils.