

Discussie

Naar een vollediger systeem van kwaliteitsbeoordeling van scholen

M.P.C. van der Werf en B.P.M Creemers
Rijksuniversiteit Groningen

Samenvatting

Op basis van gegevens van scholen over examenresultaten, zittenblijven en drop-out heeft Dronkers een overkoepelende maat geconstrueerd als index voor de kwaliteit van scholen. De validiteit van deze maat wordt in deze bijdrage ter discussie gesteld. Er wordt uiteengezet dat verschillende outputmaten, zoals numeriek rendement, efficiëntie, effectiviteit – waarin de doelen van het onderwijs worden bereikt in termen van geleerde kennis en vaardigheden – afzonderlijk beoordeeld dienen te worden, daarbij rekening houdend met achtergrond- en instroomkenmerken van leerlingen. Ook de selectiviteit van scholen, de mate waarin scholen in staat zijn een zo'n hoog mogelijk niveau te bereiken voor alle categorieën leerlingen, is een belangrijke kwaliteitsmaat. De kwaliteitsmaat van Dronkers houdt niet voldoende rekening met achtergrondkenmerken van leerlingen en contextkenmerken van scholen. De validiteit en betrouwbaarheid van de kwaliteitsmaat van Dronkers wordt getoetst door zijn resultaten te vergelijken met empirische gegevens van een grootschalig cohortonderzoek in het Voortgezet Onderwijs. Geconcludeerd wordt dat de relatie tussen de kwaliteitsmaat van Dronkers en alternatieve maten gering is. Er is meer onderzoek nodig om vast te stellen welke indicatoren van kwaliteit het meest valide zijn om de overall kwaliteit van scholen te evalueren.

1 Inleiding

Enige tijd geleden is er nogal wat commotie geweest over de gegevens van scholen die in *Trouw* zijn gepubliceerd en de rapportcijfers

van scholen die Dronkers op basis van deze gegevens had berekend. Op zichzelf is publicatie van gegevens, die als indicatie kunnen worden beschouwd voor de kwaliteit van scholen, verdedigbaar. Ouders hebben er recht op te weten of hun keuze voor een bepaalde school verantwoord is (geweest). De keuze voor één overkoepelende kwaliteitsmaat om scholen mee te beoordelen is echter niet verantwoord. Kwaliteit van scholen in termen van opbrengsten omvat meerdere aspecten die weliswaar niet los van elkaar staan, maar die ook niet zonder meer bij elkaar opgeteld kunnen worden tot een eendoordeel. Verder is in de beoordeling van de opbrengsten niet adequaat gecorrigeerd voor achtergrondkenmerken van leerlingen en scholen. Ten slotte is het van belang dat niet alleen gekeken wordt naar de opbrengsten van het onderwijs maar ook naar de wijze waarop deze opbrengsten gerealiseerd worden, dus naar de kwaliteit van het onderwijsproces in de school en in de klas.

Dit artikel handelt over de beoordeling van de opbrengsten van scholen. Op de beoordeling van de kwaliteit van processen wordt ingegaan in de conclusies. Na een korte uiteenzetting over de door Dronkers gehanteerde kwaliteitsmaat worden de bezwaren toegelicht, worden alternatieve maten voorgesteld en worden de uitkomsten van de beoordeling van scholen door Dronkers vergeleken met die van de beoordeling op andere kwaliteitsmaten.

2 Opbrengsten van scholen: rendement, efficiëntie, effectiviteit en selectiviteit

Dronkers heeft op basis van de percentages geslaagden voor het eindexamen, gemiddelde eindexamencijfers voor wiskunde, Nederlands en Engels en de percentages zittenblijvers en drop-outs in de leerjaren 2 tot en met het voorlaatste (leerjaar 5 VWO, leerjaar 4 HAVO, leerjaar 3 VBO en MAVO) een kwaliteitsmaat geconstrueerd. De score op deze maat is een gewogen optelling van de genoemde gegevens waarbij de gewichten zijn bepaald op grond van de factorladingen in een Principale Componenten Analyse. De resulterende score is vervolgens gecorrigeerd voor het percentage allochtone leerlingen door de score te vermin-

deren met het product van dit percentage en de ruwe regressiecoëfficiënt van de regressie van de score op het percentage allochtone leerlingen.

De wijze waarop de outputelementen zijn opgeteld tot een eindscore wordt door ons niet ter discussie gesteld. Bij de elementen die deel uitmaken van de eindscore kunnen echter vraagtekens worden geplaatst. De kwaliteit is in de eerste plaats beoordeeld in termen van *rendement*, het percentage geslaagden in een eindexamenjaar. Rendement van scholen is echter volgens ons meer. In productietermen gedacht is rendement het verschil tussen de opbrengsten en de kosten. In termen van opbrengsten en kosten kunnen de aantallen gediplomeerde leerlingen worden opgevat als opbrengsten en de aantallen leerlingen die uitvallen als de kosten. De aantallen uitvallers kunnen echter alleen worden bepaald als bekend is hoeveel leerlingen er in het begin waren. Rendement van onderwijs is dan dus niets anders dan het verschil tussen het aantal leerlingen dat in een bepaald jaar is ingestroomd in de eerste klas van een bepaald opleidingstype en het aantal leerlingen van die jaargang dat met een diploma de school verlaat. Om rendement goed te bepalen dient bekend te zijn hoeveel procent van de leerlingen die instromen in het eerste leerjaar van het voortgezet onderwijs met een diploma de school verlaat. Dit staat nog los van de efficiëntie. Bij *efficiëntie* gaat het om het tijdsaspect en dus om de vraag hoe lang leerlingen er over doen om het diploma te behalen. In de kwaliteitsmaat van Dronkers is efficiëntie bij benadering opgenomen, door rekening te houden met percentages zittenblijvers en drop-outs. Deze percentages zeggen echter niets over de tijd die het de eindexamenleerlingen heeft gekost om het diploma te behalen.

Rendement en efficiëntie zijn twee manieren om naar de opbrengsten van scholen te kijken. In het licht van de opvatting van onderwijs als productieproces zijn het ook consequente manieren. Maar voor de beoordeling van de kwaliteit van scholen zijn alleen rendement en efficiëntie onvolledig. Bovendien zal het hanteren van deze maten gemakkelijk kunnen leiden tot strategisch gedrag van scholen in die zin dat er sprake zal zijn van strenge selectie bij toelating en wellicht ook tot verlaging van de

prestatienormen, zowel gedurende de periode tot aan het eindexamen als ook bij het beoordelen van de examenprestaties (zie ook Bosker, Brandsma & Van Amelsvoort, 1994). Het huidige systeem van schoolonderzoeken, waarvan uit onderzoek bekend is dat de resultaten daarvan op sommige scholen grote discrepanties vertonen met de resultaten op het Centraal Schriftelijk Eindexamen (Pijl, 1991), sluit deze mogelijkheid zeker niet uit. Dit betekent niet dat eindexamenresultaten niet van belang zijn voor het bepalen van de kwaliteit van een school. Integendeel. Het bijbrengen van kennis en vaardigheden is de belangrijkste doelstelling van onderwijs en dus dient in de kwaliteitsbeoordeling betrokken te worden de mate waarin de leerdoelen zijn bereikt. De kwaliteit van onderwijs moet blijken uit wat leerlingen ervan hebben opgestoken (De Groot, 1993). Ook door de overheid wordt deze definitie van kwaliteit onderkend. In de nota 'De school als lerende organisatie' wordt gesteld dat de beoordeling van de kwaliteit van het onderwijs altijd een waardering inhoudt van de mate waarin de doelen die aan het onderwijs gesteld worden, bereikt worden (Netelenbos, 1995). De mate van doelbereiking wordt *effectiviteit* genoemd. De effectiviteit van scholen kan gemeten worden met behulp van toetsen tijdens en aan het eind van de schoolloopbaan, waarmee de prestaties van de leerlingen onafhankelijk beoordeeld worden, zodat er geen gesjoemel met de normen mogelijk is. Cito-toetsen en ook het Centraal Schriftelijk Eindexamen zijn geschikte middelen om de effectiviteit van het onderwijs te beoordelen, maar eindexamencijfers waarin zowel de resultaten van het Centraal Schriftelijk Eindexamen als die van de schoolonderzoeken zijn opgenomen zijn dat zeker niet. Als Cito-toetsscores of scores op het Centraal Schriftelijk Eindexamen echter worden gehanteerd als enige middel ter beoordeling van scholen wordt opnieuw strategisch gedrag uitgelokt, nl. een strenge selectie zowel bij toelating tot de school als gedurende 'de rit'. Alleen de beste leerlingen, die naar verwachting hoog zullen scoren op de toetsen blijven over. De rest stroomt af naar een lager onderwijstype, en zal daar dan een bijdrage leveren tot hogere scores op de toetsen, ofwel ze blijven zitten en/of raken zo gedemotiveerd dat ze de school verlaten.

Het zal duidelijk zijn dat het gebruik van alleen toetscores als kwaliteitsmaat onmiddellijk gevolgen zal hebben voor het rendement en de efficiëntie van het onderwijs, zoals eerder gedefinieerd, evenzo als alleen beoordelen op grond van rendement en efficiëntie gevolgen zal hebben voor de effectiviteit van scholen (zie ook Hofstee, 1994). De consequentie hiervan is dat zowel de *effectiviteit* als het *rendement* en de *efficiëntie* in de beoordeling van scholen moeten worden betrokken, hetgeen Dronkers op grove wijze heeft gedaan. Wat Dronkers echter over het hoofd heeft gezien is dat het hanteren van deze drie criteria en het onderbrengen ervan in één overkoepelende kwaliteitsmaat het gevaar met zich mee brengt dat scholen in hoge mate selectief zullen worden bij toelating tot bepaalde typen onderwijs en bij beslissingen over bevordering naar hogere leerjaren, opstroom en afstroom en bij de keuze van het vakkenpakket of in de toekomst van profielen. Dit is strijdig met de taken en verantwoordelijkheden van het onderwijs. Het onderwijs heeft de verantwoordelijkheid om bij alle leerlingen 'het onderste uit de kan te halen'. Dit betekent dat een school zijn uiterste best moet doen om een zo hoog mogelijke opleiding en een zo hoog mogelijk kennisniveau te realiseren voor alle leerlingen. Het begrip 'zo hoog mogelijk' betekent uiteraard niet dat voor alle leerlingen dezelfde opleiding en hetzelfde kennisniveau gerealiseerd moet worden. Er zijn nu eenmaal individuele verschillen in mogelijkheden en motivatie van leerlingen die grenzen stellen aan de mogelijkheden van een school. Met 'zo hoog mogelijk' wordt bedoeld 'gegeven de aanleg, het beginniveau en de motivatie van de leerling'. Verschillen in kwaliteit (zowel m.b.t. rendement, efficiëntie en effectiviteit) tussen scholen komen in deze opvatting tot uitdrukking in de diploma's die leerlingen behalen, de tijd die ze daarover doen en in hun scores op onafhankelijke toetsen, *rekening houdend met de aanleg, het beginniveau en het verwachte diploma* bij intrede van de leerlingen in de school. Een school waarop meer leerlingen met een HAVO-advies een HAVO-diploma en hogere prestaties halen is een betere school dan een school die met qua achtergrondkenmerken vergelijkbare leerlingen lagere opbrengsten heeft. Maar als op de laatstgenoemde school ook

meer leerlingen met een MAVO-advies en relatief lage prestaties op de Cito-eindtoets een HAVO diploma halen alsook hogere cijfers dan op de eerstgenoemde school, dan is deze school tevens minder selectief – of positief uitgedrukt: heeft een beter compenserend vermogen (zie Brandsma & Knuver, 1988). Essentieel is dus dat, ter voorkoming van een hoge mate van selectiviteit door scholen, bij de bepaling van zowel het rendement en de efficiëntie als van de effectiviteit niet alleen gekeken wordt naar de overall resultaten van een school, gecorrigeerd voor instroomkenmerken van leerlingen. Ook dienen de drie typen opbrengsten voor afzonderlijke groepen leerlingen waarvoor verschillende verwachtingen bestaan bij entree op de school afzonderlijk te worden beoordeeld. Dronkers heeft weliswaar gekeken naar de opbrengsten van scholen voor VBO-, MAVO-, HAVO- en VWO-leerlingen, maar heeft hierbij geen rekening gehouden met de verwachtingen die voor de leerlingen bestonden bij hun instroom in het eerste leerjaar. Selectiviteit van scholen dient in de beoordeling van de opbrengsten van scholen te worden verdisconteerd door de verwachtingen (of het advies) te vergelijken met het bereikte eindniveau en door rekening te houden met opstroom en afstroom van leerlingen naar hogere respectievelijk lagere onderwijstypes.

3 Probleemstelling

In het voorgaande zijn de bezwaren uiteengezet tegen de door Dronkers gehanteerde maten voor het beoordelen van de opbrengsten van scholen. De kern van de bezwaren is dat in de door Dronkers gehanteerde maten geen rekening is gehouden met de instroomkenmerken van leerlingen. Het feit dat in de cijfers wel rekening is gehouden met het percentage allochtone leerlingen kan dit probleem niet ondervangen, omdat zo'n percentage geen goede indicator is voor de individuele instroomkenmerken van de leerlingen. Het is alleen een indicator voor de leerlingpopulatie van een school waarop een school zelf, door haar selectiebeleid veel invloed kan uitoefenen. Doordat geen rekening is gehouden met de instroomkenmerken van leerlingen zijn de examenresultaten (percentages geslaagden en

eindexamencijfers) eveneens meer een uitdrukking van de selectiviteit van scholen dan van hun kwaliteit. In het hiernavolgende wordt onderzocht of de kwaliteitsindex die Dronkers heeft gehanteerd samenhangt met kwaliteitsindexen die wel zijn gecorrigeerd voor individuele leerlingkenmerken. Deze alternatieve kwaliteitsindexen volgen uit wat eerder is opgemerkt over de taken van de school: kwaliteit wordt geïndiceerd door het gerealiseerde opleidings- en kennisniveau, gegeven een bepaalde verwachting op basis van aanleg, geadviseerde opleiding en beginniveau. De kwaliteitsindexen zijn afzonderlijk berekend voor VBO-, MAVO-, HAVO- en VWO-leerlingen. In de index is dus ook rekening gehouden met de selectiviteit van de school. De vraagstellingen luiden als volgt:

- 1 *Hoe groot zijn de verschillen tussen scholen op de alternatieve kwaliteitsindexen?*
- 2 *Hoe hangen de alternatieve kwaliteitsindexen onderling samen?*
- 3 *Hoe groot is de samenhang tussen de alternatieve kwaliteitsindexen en de kwaliteitsindex van Dronkers?*

4 Methode

4.1 Steekproef

De vraagstellingen worden beantwoord met behulp van de gegevens van de scholen die meedoen aan het VOCL '93 cohortonderzoek. De gegevens van dit cohort zijn geschikter dan die van cohort '89 omdat de meeste scholen die aan het cohort '89 meewerkten in 1996/1997, het jaar waarop de rapportcijfers van Dronkers betrekking hebben, niet meer in hun oorspronkelijke gedaante bestonden vanwege opheffing of fusies. Bovendien is het instroomjaar van het VOCL '93 cohort voor VBO- en MAVO-leerlingen hetzelfde als dat van de eindexamenleerlingen in 1996/1997. Het VOCL '93 cohort heeft echter als nadeel dat er nog geen gegevens beschikbaar zijn over eindexamencijfers en diploma's. We moeten volstaan met de gegevens tot en met het derde leerjaar.

Aan het VOCL '93 cohort werken ongeveer 350 scholen voor voortgezet onderwijs mee alsmede ongeveer 20.000 leerlingen die in 1993 in het eerste leerjaar zaten. Van ongeveer 200 scholen en 13.000 leerlingen zijn gegevens

tot en met het derde leerjaar beschikbaar. In Tabel 1 staat een overzicht van het aantal scholen per onderwijstype waarvan alternatieve kwaliteitsindexen berekend konden worden en waarvan tevens rapportcijfers van Dronkers beschikbaar waren. Ook de aantallen leerlingen per onderwijstype staan in deze tabel.

Tabel 1

Aantal scholen en leerlingen in de analyses, per schooltype

	VBO	MAVO	HAVO	VWO
Aantal scholen	79	44	83	44
Aantal leerlingen	2041	1048	2189	1146

4.2 Variabelen

Voor de berekening van de kwaliteitsindexen is gebruik gemaakt van de gegevens over het gerealiseerde onderwijstype (VBO, MAVO, HAVO en VWO) en de prestaties, beide in leerjaar drie. In deze indexen zit het efficiëntie-criterium verdisconteerd doordat alleen onvertraagde leerlingen in de analyses zijn opgenomen. Per onderwijstype zijn twee kwaliteitsindexen berekend, nl. één voor de prestaties voor wiskunde en één voor de prestaties voor Nederlands. De prestaties zijn gemeten met toetsen die speciaal door het Cito zijn geconstrueerd voor de evaluatie van de Basisvorming. De toets Nederlands was voor alle leerlingen in de afzonderlijke onderwijstypen hetzelfde, de toets wiskunde bestond uit een A-versie voor MAVO-, HAVO- en VWO-leerlingen en een B-versie voor IBO/VBO-leerlingen. De toets Nederlands bestond uit 40 items en de betrouwbaarheid was .80. De A-versie van de toets wiskunde bestond uit 33 items (betrouwbaarheid .85) en de B-versie van de toets wiskunde bestond uit 33 items met een betrouwbaarheid van .89.

Als instroomkenmerken van de leerlingen zijn beschikbaar: Intelligentie, gemeten met de PSB in leerjaar 1; Sociaal milieu, geïndiceerd door het opleidingsniveau van de ouders (7 niveaus); Etniciteit (allochtoon/ niet-allochtoon); Scores op Cito entreetoetsen in het eerste leerjaar voor rekenen (betrouwbaarheid .84) en Nederlandse taal (betrouwbaarheid .76); en Geadviseerd type voortgezet onderwijs (9 niveaus).

De laatste variabele in de analyses is uiteraard

het rapportcijfer van Dronkers.

4.3 Analyses

Voor de constructie van de kwaliteitsscores is gebruik gemaakt van multilevel analyses met behulp van het programma VARCL. Na specificatie van het lege model zijn eerst de leerlingvariabelen opgenomen. Dit om na te gaan hoe groot de verschillen zijn tussen scholen na correctie voor de instroomkenmerken van de leerlingen. Deze analyses zijn gedaan per onderwijstype afzonderlijk alsmede afzonderlijk voor wiskunde en Nederlands. De uit deze analyses resulterende posterior means voor elke school zijn de alternatieve kwaliteitsscores van scholen voor Nederlands en wiskunde. Deze zijn vervolgens met elkaar gecorreleerd om na te gaan hoe groot de onderlinge samenhang tussen de alternatieve kwaliteitsscores is. Ten slotte zijn ze gecorreleerd met de rapportcijfers van Dronkers.

5 Resultaten

Allereerst worden de verschillen tussen scholen op de kwaliteitsindexen gepresenteerd. Deze zijn afzonderlijk berekend voor Nederlands en wiskunde. Dit om na te gaan of de verschillen tussen scholen voor deze vakken ongeveer even groot zijn en om na te gaan of scholen die hoog scoren op Nederlands dit ook doen voor wiskunde. De verschillen zijn afzonderlijk voor de opleidingstypen geanalyseerd (Tabel 2).

Tabel 2
Percentages onverklaarde variantie in toetsprestaties in leerjaar 3 op leerling- en schoolniveau na controle voor instroomkenmerken, per vak en schooltype

Variantie	Onderwijstype			
	VBO	MAVO	HAVO	VWO
Wiskunde				
leerling	60	56	68	68
school	12	22	14	11
Nederlands				
leerling	75	69	66	74
school	15	23	20	17

De percentages variantie op schoolniveau laten zien dat er aanzienlijke verschillen tussen scholen zijn, ook nadat rekening is gehouden met de instroomkenmerken van de leerlingen. Dit betekent dat scholen verschillen in effectiviteit, zoals afgemeten aan de prestaties van de leerlingen in het derde leerjaar op onafhankelijke toetsen. Op sommige scholen hebben leerlingen na drie jaar meer geleerd dan op andere. Dit geldt voor elk gevolgd onderwijstype, maar meer voor MAVO en HAVO en meer voor Nederlands dan voor wiskunde.

Ter beantwoording van vraagstelling 2 is per school een gemiddelde score voor Nederlands en wiskunde berekend waarin rekening is gehouden met de instroomkenmerken van de leerlingen (de zogenaamde posterior means). Dit zijn de kwaliteitsscores van de scholen voor Nederlands en wiskunde. Vervolgens zijn correlaties berekend tussen deze twee kwaliteitsscores. In Tabel 3 staan de resultaten.

Tabel 3
Correlatie tussen de gemiddelde scores van Nederlands en wiskunde in leerjaar 3, per schooltype

	VBO	MAVO	HAVO	VWO
	.45	.34	.24	.24

Uit deze tabel blijkt dat de correlaties niet erg hoog zijn. Scholen die hoog scoren voor wiskunde scoren niet tevens hoog voor Nederlands. De correlatie is het hoogst voor het VBO. Een verklaring hiervoor kan zijn dat in het VBO vaak door een en dezelfde docent zowel wiskunde als Nederlands wordt gegeven. Docenten in het VBO zijn vaak meervoudig bevoegd op basis van oude LO aktes (Den Boer, Kuyper & Rekveld, 1996). In de andere schooltypen is meestal sprake van vakdocenten met een tweede- of eerstegraads bevoegdheid.

De derde vraagstelling heeft betrekking op de vergelijking van de scores van de scholen op de alternatieve kwaliteitsmaten en de rapportcijfers van Dronkers. Voor het maken van deze vergelijking zijn de rapportcijfers van Dronkers van de scholen die participeren aan VOCL '93 gecorreleerd met de posterior means voor Nederlands en wiskunde van die scholen. De resultaten staan in Tabel 4.

Tabel 4

Correlatie tussen de gemiddelde scores van Nederlands en wiskunde in leerjaar 3 en rapportcijfers Dronkers, per schooltype

	Rapportcijfer			
	VBO	MAVO	HAVO	VWO
Nederlands	-.13	.10	.06	-.04
Wiskunde	-.08	.24	.41	.05

Uit de correlaties blijkt dat de samenhang tussen de rapportcijfers van Dronkers en de kwaliteitscores op basis van toetsprestaties in het derde leerjaar over het algemeen zeer laag en soms zelfs negatief is. De samenhang met de rapportcijfers is het hoogst voor de wiskunde-prestaties in HAVO-klassen, maar ook deze correlatie is veel lager dan men zou mogen verwachten. In het algemeen mag geconcludeerd worden dat de kwaliteitsbeoordeling op basis van de overall kwaliteitscore niet samenhangt met andere kwaliteitsindexen die aangeven wat onvertraagde leerlingen in een bepaald onderwijstype aan kennis van bepaalde vakken hebben opgedaan aan het eind van het derde leerjaar, rekening houdend met hun beginniveau, sociale en etnische achtergrond en advies. In de probleemstelling is geopperd dat dit te maken heeft met het feit dat Dronkers niet adequaat heeft gecorrigeerd voor instroomkenmerken van de leerlingen. Deze verklaring wordt sterker indien de rapportcijfers van Dronkers beter zouden correleren met de ongecorrigeerde scores op de alternatieve kwaliteitsmaten, dat wil zeggen de gemiddelde toetsprestaties waarin geen rekening is gehouden met de instroomkenmerken van de leerlingen. Analyse van de correlaties met de ongecorrigeerde gemiddelden (niet in tabel) laat echter zien dat deze in het algemeen weliswaar iets hoger zijn, maar nog steeds niet erg hoog. De hoogste correlatie is nu .45, wederom voor wiskunde in HAVO.

In Tabel 5 staan de correlaties tussen de ongecorrigeerde gemiddelde eindexamencijfers, zoals in *Trouw* gepubliceerd en de ongecorrigeerde alternatieve kwaliteitscores. Voor de volledigheid zijn in deze tabel ook de correlaties opgenomen tussen de rapportcijfers van Dronkers en de gemiddelde eindexamencijfers.

Tabel 5

Correlatie tussen de gemiddelde scores van Nederlands en wiskunde in leerjaar 3 en gemiddelde eindexamencijfers van Dronkers en tussen de rapportcijfers van Dronkers en gemiddelde eindexamencijfers, per schooltype

	Gemiddelde eindexamencijfer			
	VBO	MAVO	HAVO	VWO
Nederlands	.30	.23	-.08	.08
Wiskunde	.14	.06	.05	.30
Rapportcijfer	-.16	.79	.07	.34

De correlaties tussen de ongecorrigeerde gemiddelden voor Nederlands en wiskunde en de ongecorrigeerde eindexamencijfers zoals gepubliceerd in *Trouw* zijn laag tot zeer laag. Het feit dat ook de correlaties tussen eindexamencijfers en de rapportcijfers van Dronkers laag zijn (behalve in MAVO) en in VBO zelfs negatief, terwijl de rapportcijfers mede gebaseerd zijn op de eindexamencijfers is zeer vreemd. Het doet vermoeden dat de constructie van de rapportcijfers op basis van een gewogen optelling van eindexamencijfers, rendement en zittenblijven/uitval geen valide manier is om een overall index voor kwaliteit te construeren, althans niet zoals door Dronkers is verricht. Overigens blijkt dit ook reeds uit de resultaten van de factoranalyses van Dronkers zelf.

6 Conclusies

Met de analyses zoals hierboven gepresenteerd is aangetoond dat het beoordelen van scholen op basis van een overkoepelende kwaliteitsindex een riskante aangelegenheid is, zeker wanneer daarin onvoldoende rekening wordt gehouden met de instroomkenmerken van de leerlingen. De kwaliteitsindex van Dronkers hangt nauwelijks samen met andere indexen van kwaliteit, zowel indien deze worden gecorrigeerd voor instroomkenmerken van leerlingen als indien deze ongecorrigeerd worden gehanteerd. De overkoepelende index hangt zelfs nauwelijks samen met een afzonderlijke index die van de overkoepelende index deel uitmaakt. Uit deze resultaten kan geconcludeerd worden dat de validiteit van de beoordeling van scholen door Dronkers zeer twijfelachtig is. Dit ligt niet alleen aan het feit dat er slechts grove maten beschikbaar waren, maar

ook aan de wijze waarop de overkoepelende index is geconstrueerd.

Dit leidt tot de vraag of het eigenlijk wel mogelijk is een overkoepelende index te construeren. Voorlopig denken we dat het antwoord hierop negatief moet zijn. In de eerste plaats is duidelijk gemaakt dat het bij de kwaliteit van scholen gaat om meerdere aspecten, zoals rendement, efficiëntie, effectiviteit en selectiviteit, die kunnen worden beschouwd als verschillende dimensies van kwaliteit. Zolang onduidelijk is hoe die dimensies zich tot elkaar verhouden, is een optelling van de scores van scholen op deze dimensies – gewogen dan wel ongewogen – onmogelijk. Dit geldt ook voor het optellen en middelen van eindexamenvakken tot gemiddelde eindexamenscores. In de gepresenteerde analyses hebben we laten zien dat scholen weinig consistent zijn in hun output over vakken heen. Een school die hoog scoort voor wiskunde kan tegelijkertijd laag scoren voor Nederlands. Een kwaliteitsindex kan dus niet volstaan met scores voor slechts enkele vakken en bovendien zullen de vakken gewogen moeten worden bij het berekenen van een gemiddelde. In plaats van een empirische weging op grond van factorladingen zoals Dronkers heeft gedaan zal hierbij gekeken moeten worden naar de zwaarte van de vakken en naar de voorspellende waarde van de vakken voor verdere opleiding of studie. Op dit moment is hierover weinig bekend.

In de tweede plaats is ook bekend dat scholen weinig stabiel zijn in hun opbrengsten over schooljaren heen. Een school die het ene jaar hoog scoort op een kwaliteitscriterium hoeft een jaar later niet per definitie ook hoog te scoren. Criteria die meer stabiel zijn over de tijd zouden misschien in de algemene index zwaarder moeten wegen dan minder stabiele criteria. Een gevaar hiervan is echter dat wellicht de minder essentiële criteria het meest stabiel blijven te zijn.

Afgezien van bovengenoemde problemen is het echter voor een eerlijke vergelijking tussen scholen ook van belang dat correctiefactoren worden ingebouwd waarvan bekend is dat die in belangrijke mate verschillen in output verklaren maar die los staan van de invloed van de school. Aan de andere kant moet niet gecorrigeerd worden voor factoren die duidelijk geen invloed hebben. In dit verband wijzen we erop

dat in onze analyses de etniciteit van leerlingen er niet toe doet voor de verklaring van verschillen in prestaties tussen leerlingen in het VWO, nadat reeds rekening is gehouden met SES, intelligentie en beginniveau. Bij de berekening van kwaliteitsscores hoeft voor VWO scholen dan ook geen rekening te worden gehouden met de etniciteit van de leerlingen. Correctie voor het percentage allochtonen op schoolniveau is alleen nodig als dit, na correctie voor etniciteit en andere instroomkenmerken op leerlingniveau nog steeds een zelfstandige bijdrage levert aan de verschillen in output tussen scholen. Het zonder meer corrigeren voor het percentage allochtonen op schoolniveau zoals Dronkers dat heeft gedaan zou wel eens kunnen hebben geleid tot een onterecht hoge waardering van VWO-afdelingen in een scholengemeenschap met veel allochtone leerlingen en een onterecht lage waardering van scholen met weinig van deze leerlingen. Voor de goede orde: ook wij hebben niet de pretentie in onze analyses afdoende te hebben gecorrigeerd voor instroom- en achtergrondkenmerken van de leerlingen. Het bewijs daarvoor is dat er nog zeer veel variantie op leerling- en schoolniveau onverklaard is. We vermoeden dat veel van de variantie op leerlingniveau verklaard kan worden door de motivatie en inzet van de leerlingen. Nadere analyses zullen dit moeten uitwijzen. Daarnaast zullen er vermoedelijk op schoolniveau en in de context van de school factoren zijn die niets te maken hebben met de kwaliteit van het onderwijsleerproces van de school maar die wel deels de verschillen tussen scholen kunnen verklaren. Als deze factoren er zijn zal daarmee in de kwaliteitsindex ook rekening moeten worden gehouden. Te denken valt bijvoorbeeld aan de buurt of regio waarin de school staat, opleiding en leeftijd van leraren en dergelijke.

Samenvattend kan geconcludeerd worden dat het thans niet mogelijk is om scholen te beoordelen op basis van *een* kwaliteitsindex. Dat neemt niet weg dat wel resultaten van scholen op beschikbare opbrengstmaten kunnen worden gepubliceerd, zoals bijvoorbeeld percentages geslaagden, eindexamencijfers en dergelijke, overeenkomstig de bestaande plannen tot het uitgeven van 'kwaliteitskaarten' door de inspectie. Voorlopig zullen daarop alleen nog cijfers omtrent rendement en exa-

menresultaten komen te staan en vanaf 1999 ook doorstroom naar en resultaten in het vervolgonderwijs. Op zich is daar niets op tegen, onder de conditie dat het bij eindexamenresultaten gaat om de cijfers op het Centraal Schriftelijk Eindexamen. Het moet echter duidelijk zijn dat hiermee niet een soort keurmerk aan scholen kan worden gegeven op grond waarvan De Kwaliteit van de school is vastgesteld. De reactie van Dronkers (1997) op de kwaliteitskaarten onderschrijven we dus niet waar hij stelt dat de gegevens daarvan onvoldoende zijn voor ouders omdat een samenvattend oordeel ontbreekt. Ouders zullen volgens Dronkers door de bomen het bos niet meer zien. Het verdient echter de voorkeur dat ouders op basis van meerdere gegevens zelf uitmaken waar 'het bos' is dan dat hun een 'bos' wordt beloofd op basis van een invalide kwaliteitsmaat. Dit betekent niet dat een overall kwaliteitsmaat per definitie onmogelijk is. Om dit te kunnen bepalen is nader onderzoek nodig in de vorm van analyses op bestaand materiaal. De gegevens die de inspectie in haar bezit heeft zijn hiervoor eigenlijk te beperkt gebleken. Momenteel zijn er wel grote databestanden beschikbaar die van twee cohorten leerlingen op ruim 400 scholen voor voortgezet onderwijs gegevens bevatten over hun schoolloopbaan. Op basis van deze gegevens kunnen de opbrengsten van scholen worden berekend voor diverse criteria, zoals behaalde diploma's, bereikt kennisniveau in leerjaar 3 en op het moment van eindexamen (eindexamencijfers), opstroom, afstroom en zittenblijven en kán worden gecorrigeerd voor instroom en achtergrondkenmerken van leerlingen en contextkenmerken van de school. Een mooi opbrengstcriterium, waarin zowel de dimensie rendement als de dimensie efficiëntie is verdisconteerd, is de score op de leerjarenladder (Bosker, Van der Velden & Hofman, 1985). De resultaten van scholen op deze afzonderlijke opbrengstcriteria kunnen onderling worden vergeleken evenals de resultaten over twee cohorten heen als indicatie voor de stabiliteit. Op basis van deze analyses kunnen diverse kwaliteitsindexen worden geconstrueerd en met elkaar vergeleken. De indexen die het meest met elkaar samenhangen en het meest stabiel zijn over de tijd kunnen wellicht uiteindelijk deel gaan uitmaken van een overkoepelende kwaliteitsbeoordeling van scholen.

Zoals in de inleiding echter reeds is opgemerkt is het beoordelen van de kwaliteit van scholen in alleen termen van opbrengsten nogal beperkt. Het gaat er ook om hoe die opbrengsten worden bereikt. In haar reactie op het *Trouw*-onderzoek geeft de Staatssecretaris te kennen dat kwaliteitsoordelen zo veel mogelijk over de volle breedte dienen te worden bepaald (Netelenbos, 1997). Zeker indien kwaliteitsoordelen voor ouders bedoeld zijn, is eenzijdige nadruk op output van scholen onvoldoende. Ouders zijn niet alleen geïnteresseerd in de resultaten van scholen maar ook of hun kinderen het op school naar hun zin (zullen) hebben (Mertens, Van Dommelen & De Rijcke, 1994). Ook leerlingen zelf hebben daar baat bij. Uit onderzoek naar de ontwikkeling van motivatie van leerlingen in de eerste drie jaar van het voortgezet onderwijs blijkt dat er een positief verband is tussen de mate waarin leerlingen het leuk vinden op school en waardering hebben voor de leraren enerzijds en motivatie en prestaties anderzijds (Kuyper & Swint, 1996). Voor de beoordeling van de kwaliteit van scholen dient daarom ook gekeken te worden naar de kwaliteit van de schoolorganisatie, het schoolklimaat en de kwaliteit van het instructieproces in de klas. Het ligt voor de hand dat op scholen die hoger scoren op deze proceskenmerken leerlingen met meer plezier naar school gaan, een positievere houding tegenover leren hebben, meer gemotiveerd zijn en dergelijke en dat leraren met meer enthousiasme lesgeven, meer hun vak bijhouden, de leertijd effectiever gebruiken en dus minder huiswerk hoeven te geven. Uitgezocht zal moeten worden of scholen op deze kenmerken verschillen en of dit samenhangt met hun opbrengsten. Met nadruk merken we op, aansluitend bij Hofstee (1992), dat het hierbij alleen dient te gaan om procesindicatoren die voorspellend zijn voor de output en waarop scholen zelf invloed kunnen uitoefenen. Scholen zijn zelf immers het meest verantwoordelijk voor de kwaliteit van het onderwijsproces en de output daarvan (Mertens, 1995). Tot dusver heeft empirisch onderzoek nog maar weinig procesvariabelen opgeleverd die substantieel en consistent samenhangen met de output van scholen. De meeste van deze variabelen hebben te maken met tijd en gelegenheid tot leren (Cremers, 1995). Vooralsnog pleit dit voor een

beperkte set van indicatoren voor de beoordeling van de kwaliteit van het onderwijsproces op scholen. Wellicht leidt dit ook tot opruiming van een groot aantal procesindicatoren die deel uitmaken van de thans in zwang zijnde instrumenten voor zelfevaluatie door scholen (Van de Grift, 1996). Veel van de variabelen in deze instrumenten hebben met de kwaliteit van het onderwijs in termen van output niets van doen en zijn derhalve niet van belang in het kader van de kwaliteitszorg. Verder onderzoek zal moeten uitwijzen welke procesvariabelen wel relevant zijn om in de toekomst in een beoordelingssysteem op te nemen.

Literatuur

- Boer, P. den, Kuyper, H., & Rekveld, I. (1996). *Vakcombinaties in de Basisvorming*. Groningen: GION.
- Bosker, R.J., Brandsma, J.T., & Amelvoort, H.W.C.H., van. (1994). Kwaliteitsverbetering door output-financiering? In B.P.M. Creemers (red.), *Deregulering en de kwaliteit van het onderwijs* (pp. 149-161). Groningen: RION monografieën 17.
- Bosker, R.J., Velden, R.K.W. van der, & Hofman, W.H.A. (1985). *Een generatie geselecteerd. Deel 1: de schoolloopbanen*. Groningen: RION.
- Brandsma, H.P., & Knuver, J.W.M. (1988). Organisatorische verschillen tussen basisscholen en hun effect op leerlingprestaties. *Tijdschrift voor Onderwijs Research*, 13, 201-222.
- Creemers, B. (1995). Process indicators on school functioning and the generalisability of school factor models across countries. In *Measuring the quality of schools* (pp. 105-118). Parijs: OECD.
- Dronkers, J. (1997). Netelenbos komt met volstrekt onvoldoende kwaliteitskaart van scholen. *Trouw*, 31 december 1997, p.11.
- Grift, W. van de (1996). Autonomie, deregulering en kwaliteit. *Basisschool Management*, mei 1996, 12-17.
- Groot, A.D. de (1993). *Denken over onderwijs. Analyses en kritieken van A.D. de Groot*. Den Haag: SVO.
- Hofstee, W.K.B. (1992). Onderwijsindicatoren: seizoensopruiming. In B.P.M. Creemers & W.Th.J.G. Hoeben (red.), *Indicatoren van onderwijs-effectiviteit* (pp. 77-81). Groningen: ICO.
- Hofstee, W.K.B. (1994). Hoe bestuurbaar is onderwijs? In B.P.M. Creemers (red.), *Deregulering en de kwaliteit van het onderwijs* (pp. 59-67). Groningen: RION monografieën 17.
- Kuyper, H., & Swint, F.E. (1996). *Microscopisch schoolloopbaanonderzoek. De eerste drie jaren in het voortgezet onderwijs*. Groningen: GION.
- Mertens, F. (1995). De school van de toekomst zorgt zelf voor kwaliteit. In P. Appelfhof & F. Mertens, *Eenheid strategisch beleid* (pp.27-32). Zoetermeer: Ministerie van OC&W.
- Mertens, F., Dommelen, J. van, & Rijke, F.J.M. (1994). Onderwijskwaliteit als recht en als opdracht. In B.P.M. Creemers (red.), *Deregulering en de kwaliteit van het onderwijs* (pp. 1-13). Groningen: RION monografieën 17.
- Netelenbos, T. (1995). *De school als lerende organisatie. Kwaliteitsbeleid op scholen voor primair en voortgezet onderwijs*. Zoetermeer: Ministerie van OC&W.
- Netelenbos, T. (1997). Kwaliteit. *Uitleg*, 26 (5), p.97.
- Pijl, Y.J. (1991). *Discrepancies tussen het schoolonderzoek en het Centraal Schriftelijk Eindexamen*. Groningen: RION.

Auteurs

M.P.C. van der Werf is senior onderzoeker bij het GION en coördinator van de VOCL cohort studies in het voortgezet onderwijs.

B.P.M. Creemers is hoogleraar bij de vakgroep Onderwijskunde van de Rijksuniversiteit Groningen en directeur van het GION.

Adres: GION, Rijksuniversiteit Groningen. postbus 1286, 9701 BG Groningen.
E-mail m.p.c.van.der.werf@ppsv.rug.nl

Abstract

Towards a more complete system of quality assessment of schools

M.P.C. van der Werf & B.P.M. Creemers. *Pedagogische Studiën*, 1999, 76, 51-60.

On basis of school data about examination results, retention rate and drop-out rate one output measure was constructed by Dronkers, indicating the quality of schools. The validity and reliability of this quality indicator is questionable. It is made clear

that different output measures, such as numerical profit, efficiency, effectiveness – the degree of goal attainment in terms of learned knowledge and skills – have to be evaluated separately, taking into account student background and intake characteristics. Also the selectivity of schools, the degree to which schools are able to reach the highest as possible educational level for all categories of students, given their intellectual abilities and prior knowledge, is an important quality indicator. The quality indicator of Dronkers does not take into account sufficiently the student background characteristics and school context variables. The validity and reliability of the quality indicator of Dronkers is tested by comparing his results with empirical data of a large scale cohort study in secondary education. It is concluded that the relationship between the quality indicator of Dronkers and alternative indicators is very low. More research is needed to establish which indicators are the most valid and reliable to evaluate the overall quality of schools.

Discussie

Openbare prestatiegegevens van scholen; nuttigheid en validiteit

R.J. Bosker en J. Scheerens

Universiteit Twente

Samenvatting

Het publiceren van de opbrengsten van scholen voor voortgezet onderwijs wordt belicht vanuit drie toepassingscontexten en een conceptuele en technische beschouwing over wat toegevoegde waarde nu precies is.

De toepassingscontexten die besproken worden zijn: ondersteuning van schoolkeuzeprocessen, het afleggen van rekenschap (accountability) door scholen en schoolverbetering. Vanwege de relatief beperkte omvang van de verschillen tussen de scholen, onvolledige kennis over wat nu de kern is van de onderwijsproductiefunctie, en de beperkte besturingsmogelijkheden van hogere besturingsniveaus in het onderwijs, wordt beargumenteerd dat het publiceren van opbrengstgegevens van scholen slechts van beperkte waarde is voor elk der toepassingscontexten.

In de discussie over het begrip toegevoegde waarde en zijn meting, wordt betoogd dat de definitie in elk geval als uitgangspunt dient te hebben, de mate waarin individuele leerlingen boven of beneden de (statistische) verwachting presteren. Ouders hebben bovendien een ander soort toegevoegde waarde op het oog als ze een school kiezen dan andere gebruikers van de opbrengstinformatie: ouders willen weten wat de school voor hun kan betekenen, anderen willen een zuivere maat voor de inspanning die de school als zodanig geleverd heeft. Daar komt bij dat ouders met name geïnteresseerd zijn in differentiële toegevoegde waarde: de specifieke betekenis van de school voor het eigen kind gegeven diens eigenschappen. Het artikel besluit met een uiteenzetting over de mogelijke negatieve gevolgen van het openbaar maken van schoolresultaten, maar de slotsom is dat de

praktijk als zodanig vermoedelijk opbrengstverhogend zal werken.

1 Inleiding

De tweede *Trouw*-publicatie (in juni 1998) over de opbrengsten van scholen voor voortgezet onderwijs lokte aanmerkelijk minder discussie uit dan de eerste, die in het najaar van 1997 verscheen. Tot nu toe waren de twee belangrijkste elementen in de discussie of het wel verstandig was opbrengstgegevens van scholen openbaar te maken en of de door Dronkers gehanteerde techniek voor het becijferen van scholen wel adequaat was (Veenstra, Dijkstra, Peschar & Snijders, 1998; Van der Werf & Creemers, 1999; Van Putten & Van der Kamp, 1998; Dronkers, 1998). Langzamerhand verstormt de discussie over het eerste element en ten aanzien van het tweede element is een Babylonische spraakverwarring rond het begrip 'toegevoegde waarde' opgetreden. In de onderhavige bijdrage zullen we onze standpunten ter zake uiteenzetten.

Hoewel de *Trouw*-publicatie en ook de kwaliteitskaarten van scholen die door de inspectie van het onderwijs uitgegeven worden primair bedoeld zijn om schoolkeuzeprocessen te ondersteunen, zijn er nog twee andere beweegredenen voor het openbaar maken van prestatiegegevens van scholen, namelijk enerzijds zelfevaluatie en schoolverbetering en anderzijds externe controle op de kwaliteit van het functioneren van scholen, een toepassingscontext die wordt aangeduid met de term 'accountability'. We zullen eerst op de diverse toepassingscontexten ingaan, vervolgens de meer technische aspecten van het bepalen van de door scholen gerealiseerde opbrengsten behandelen en ten slotte beide constituerende elementen van deze bijdrage met elkaar verbinden. Slechts zijdelings, en alleen daar waar het de betoogtrant niet stoort, zullen we op de technische onvolkomenheden van de werkwijze van Dronkers ingaan.

2 De toepassingscontext

2.1 Hoe ouders kiezen en wat er te kiezen valt

Uit onderzoek van Derriks, De Kat en Deckers (1995) blijkt dat in Nederland drie motieven een belangrijke rol spelen bij de schoolkeuze: kwaliteit, afstand en denominatie. Prestatiegegevens worden overigens niet genoemd als kwaliteitsindices, maar wel, bijvoorbeeld, de studiebegeleiding, de grootte van de school en het schoolregime. Dit beeld komt overeen met de resultaten van een casestudy-onderzoek van Lam (1998), die met 20 ouders interviews voerde met als vraag welke informatie ze graag wilden hebben ter ondersteuning van de schoolkeuze. De resultaten sporen met die van Derriks c.s.: procesindicatoren als pedagogisch klimaat, de sfeer op de school, didactische werkvormen, klassengrootte, aansluiting op het vervolgonderwijs e.d. worden belangrijker gevonden dan opbrengstgegevens.

Als ouders zich zouden laten leiden door opbrengstgegevens van scholen bij hun schoolkeuze, hebben ze dan inderdaad iets te kiezen? De rapportcijfers die Dronkers in de eerste *Trouw*-publicatie uitdeelde suggereren een enorme kwaliteitsvariatie, maar dat is natuur-

lijk een artefact van de door hem gehanteerde methodiek, waarbij per definitie altijd enen en tien worden uitgedeeld¹. Hebben ouders veel te kiezen? In Tabel 1 geven we een overzicht van het aantal scholen per schoolsoort dat zich binnen een gemeentegrens bevindt.

De meerderheid der gemeenten biedt per schoolsoort aan ouders geen alternatief: er is maar één school. Ruime keuzemogelijkheden (3 of meer scholen), althans als we voorbij gaan aan keuzes die zich langs denominatieve lijnen voltrekken, zijn slechts in ca. 20% van de gemeenten aanwezig. Hebben ouders dan ook nog iets te kiezen, in de zin dat de scholen die in aanmerking komen verschillen in hun opbrengsten? Tabel 2 bevat daartoe gegevens van enkele kengetallen die de inspectie op de zogenaamde kwaliteitskaarten van scholen vermeldt (Inspectie van het onderwijs, 1998).

Met de interkwartiele range geven we het verschil weer tussen de scholen die tot de onderste 25% slechtst presterende behoren en de scholen die tot de groep der 25% best presterende scholen behoren. Ouders en hun kinderen die binnen het agrarisch onderwijs een school moeten kiezen hebben feitelijk geen keuzemogelijkheden. De meeste variatie in gemiddelde examencijfers vinden we in het

Tabel 1
Aantal scholen per schoolsoort per gemeente

	VBO	MAVO	HAVO	VWO
1 school	64%	56%	56%	54%
2 scholen	19%	22%	20%	20%
3 scholen	9%	9%	10%	10%
4 scholen	3%	4%	6%	6%
5 of meer scholen	5%	9%	8%	8%
totaal aantal gemeenten	246	340	216	214

(bron: Bosker, 1998)

Tabel 2
Variatie in opbrengsten tussen scholen binnen gemeenten; geschatte interkwartiele range

	VBO	MAVO	HAVO	VWO
Onvertraagd in de bovenbouw	11%	6%	14%	16%
Gemiddeld cijfer centraal schriftelijk examen	n.v.t.	0.22	0.17	0.29
% vakken op C-niveau				
Techniek	24%	n.v.t.	n.v.t.	n.v.t.
Zorg & welzijn	28%	n.v.t.	n.v.t.	n.v.t.
Economie	28%	n.v.t.	n.v.t.	n.v.t.
Landbouw	0%	n.v.t.	n.v.t.	n.v.t.

(bron: Bosker, 1998)

VWO met verschillen van 6.36 tegenover 6.65, terwijl het bij het HAVO gaat om 6.25 tegenover 6.42. Dat zijn uiterst smalle marges. In het AVO vinden we de ruimste marges qua onvertraagde doorstroom wederom in het VWO met 50% tegenover 66%, met als betekenis dat op de 25% laagst scorende VWO-scholen slechts 1 op de 2 leerlingen zonder zittenblijven vanaf leerjaar 3 het VWO-diploma behaalt, terwijl dit op de 25% hoogst scorende scholen 2 op de 3 leerlingen betreft. Dat zijn relevante verschillen die bovendien ook in het HAVO gevonden worden, maar in het MAVO gaat om een zeer geringe bandbreedte met 83 tegenover 89%. Grote variaties doen zich voor in het VBO waar het de keuze van vakken op C-niveau betreft. Bij de sector 'zorg & welzijn', bijvoorbeeld, gaat het om een geschat verschil van 38 tegenover 66%.

Drinkers roert zelf het probleem van de normering aan: wanneer presteert een school nu onvoldoende en wanneer voldoende? We zouden als standaard kunnen stellen dat een school onvoldoende opbrengsten heeft als de rendementsgegevens onder de 50% liggen (meer dan 1 op de 2 leerlingen blijft in de bovenbouw in één der leerjaren zitten en voor het VBO bovendien: minder dan 1 op de 2 leerlingen slaagt met minimaal 3 vakken op C-niveau), of als de gemiddelde examencijfers onder de 6 liggen. We zouden ook een standaard kunnen formuleren waarboven scholen ruim voldoende presteren. Voor de rendementsgegevens zouden we die op 90% kunnen stellen, en voor de examencijfers op een 7 of hoger gemiddeld. Wat valt er dan te kiezen c.q. te vermijden (zie Tabel 3)?

Met de, overigens vrij willekeurig geformuleerde, standaarden in de hand kunnen ouders en hun kinderen bij hun keuze in elk geval binnen het AVO de minder presterende scholen vermijden. In het VBO is dat welhaast onmogelijk, althans als we kijken naar het criterium dat minstens de helft der leerlingen het eind-examen aflegt met minimaal 3 vakken op het C-niveau. Op zoek naar excellentie zullen velen echter ontdekken dat er in hun omgeving geen school van hun gading te vinden is.

We zien dat het beeld nogal divers is, en in sommige gevallen is er dus iets te kiezen, en in andere gevallen leidt het publiceren van de opbrengstgegevens van scholen tot de geruststelling dat elke keuze (binnen de mogelijkheden, en alleen dan als er iets te kiezen is) even goed of even slecht is. In sommige gevallen hebben ouders en hun kinderen alleen een serieus alternatief als ze bereid zijn ook buiten de eigen gemeentegrenzen te zoeken. Dit laatste speelt inderdaad een rol in de Verenigde Staten van Amerika en in Engeland, waar gezinnen die om economische redenen moeten verhuizen, zich soms bij het zoeken van een huis laten leiden door het gegeven of er in de buurt een school gevestigd is die er in de 'rankings' goed voor staat (Lam, 1998).

2.2 Het 'accountability' perspectief

Het openbaar maken van de resultaten van scholen wordt ook verdedigd vanuit 'accountability' overwegingen, namelijk dat scholen rekenschap dienen af te leggen over met welk succes de gemeenschapsgelden zijn besteed. Dat rekenschap afleggen geschiedt dan zowel

Tabel 3
'Extreem' scorende scholen

	rendement bovenbouw		examencijfer	
	<50%	>90%	<6	>7
VBO	0.2%	30.0%	n.v.t.	n.v.t.
MAVO	0.8%	28.1%	5.2%	3.4%
HAVO	33.7%	0.4%	7.5%	0.2%
VWO	27.0%	0.2%	4.4%	2.4%
3 of meer vakken op C-niveau				
Techniek	84.5%	1.3%	n.v.t.	n.v.t.
Zorg & welzijn	48.3%	5.5%	n.v.t.	n.v.t.
Economie	18.8%	24.8%	n.v.t.	n.v.t.
Landbouw	71.9%	3.1%	n.v.t.	n.v.t.

(bron: Bosker, 1998)

naar hogere bestuursorganen, als naar de zittende ouders en de lokale gemeenschap.

Veel staten in de VS hebben hun 'school report cards' ingevoerd als onderdeel van beleid gericht op accountability van schooldistricten. Het afleggen van rekenschap door scholen past bij een andere besturingsfilosofie van het onderwijs, waarin de overheid meer afstand bewaart en scholen autonomer kunnen opereren. De overheid kan zich vervolgens beperken tot het stellen van de doelen ('management by objectives'). Vervolgens kan ze nagaan of de scholen die doelen inderdaad bereiken en in voorkomende gevallen sancties opleggen. In het verleden is in Nederland met die gedachte gespeeld toen de overheid overwoog een vorm van outputfinanciering in het voortgezet onderwijs in te voeren (Van Amelsvoort, Bosker & Brandsma, 1993). Het publiceren van opbrengstgegevens van scholen is een alternatief voor outputfinanciering. De beloning of sanctie wordt nu door de marktwerking geregeld: goed presterende scholen trekken meer leerlingen aan, slecht presterende scholen zullen uiteindelijk door een gebrek aan aanmeldingen de tent moeten sluiten.

De vraag in hoeverre besturing vanuit hogere niveaus als een toepassingscontext van belang is, hangt samen met de autonomie van scholen. In Tabel 4, ontleend aan een onderzoek dat is uitgevoerd in het kader van het OECD onderwijsindicatorenproject (Scheerens, 1998) wordt aangegeven welk percentage van beslissingen in een bepaald domein, door de school zelf, of door hogere bestuursorganen wordt genomen.

Tabel 4
Autonome besluitvorming door scholen in het voortgezet onderwijs in vijf landen (als proportie beslissingen genomen door de school versus door hogere bestuursorganen)

	autonoom binnen kader	volledig autonoom
Denemarken	31	10
Duitsland	37	11
Engeland	60	40
Frankrijk	29	9
Nederland	75	39

(bron: Scheerens, 1998)

In deze tabel zijn de gegevens van vijf landen opgenomen: Denemarken en Duitsland, waar geen enkele vorm van openbaar maken van schoolresultaten bestaat, Engeland en Frankrijk, landen die een lange traditie op dit gebied hebben en ten slotte Nederland, waar dit sinds vorig jaar gebeurt. Er is een onderscheid gemaakt tussen beslissingen die scholen kunnen nemen binnen door hogere bestuursorganen vastgestelde kaders en beslissingen die scholen volledig autonoom kunnen nemen. In vergelijking met andere landen zijn Nederlandse scholen voor voortgezet onderwijs zeer autonoom. Opmerkelijk is overigens ook dat in Denemarken en Duitsland, waar geen publicatie van opbrengstgegevens plaats vindt, scholen minder autonoom zijn dan in Engeland en Nederland, waar dat wel gebeurt. Merkwaardig is de positie van Frankrijk, waar scholen nauwelijks autonoom beslissingen kunnen nemen, maar waar desalniettemin over de opbrengsten van scholen ruim gepubliceerd wordt.

Ervaringen in landen waarin het publiceren van opbrengstgegevens duidelijk vanuit een accountability-perspectief plaatsvindt, wijzen overigens op een nogal beperkte daadwerkelijke benutting. Zo komen Cibulka en Derlin (1995, pp. 493-503) tot de conclusie dat "SPR (school performance reporting; rjb en js) is not considered very important by policy-makers or the general public", en is nog lang niet aangehouden dat "SPR can become a potent, effective policy lever". Wel zien zij het als één van de krachtigste symbolen waar de overheid mee voor de dag kan komen om aan het publiek duidelijk te maken dat de beleidsmakers erg hun best doen om de kwaliteit van het onderwijs te verbeteren.

2.3 De marges voor schoolverbetering

De derde toepassingscontext ligt in het gebruik van de gegevens door scholen zelf, in het kader van zelfevaluatie en schoolverbetering. Of dit een realistische toepassingscontext is, hangt af van de mate van 'maakbaarheid' van de opbrengsten door de scholen. Een eerste inzicht in dit vraagstuk kunnen we ontlennen aan onderzoek van Luyten (1994), die examengegevens van een aantal jaren analyseerde om de potentiële bijdrage van scholen in dezen af te schatten (zie Tabel 5).

Tabel 5
De compositie van variantie in examenprestaties

	variantiebronnen	percentage
leerling	85	
school	15	
- per vak		40
- per jaar		8
- per vak per jaar		27
- over vakken en jaren heen		25

(bron: Luyten, 1994)

Prestatieverschillen tussen leerlingen blijken voor 15% te herleiden te zijn op de school die ze bezoeken. Nadere analyse leert echter dat de stabiele (over de jaren heen) en consistente (over de examenvakken heen) schoolbijdrage echter slechts een kwart van die 15% uitmaakt. Er treden enorme variaties tussen de vakken op (een school die het bijvoorbeeld goed doet voor de exacte vakken hoeft het niet ook goed te doen voor de moderne vreemde talen), per jaar zijn de examenresultaten weer anders, en bovendien wisselen de resultaten per jaar per vak. De marges voor algehele schoolverbetering zijn dus beperkt en het lijkt belangrijker

Tabel 7
De impact (als correlaties) van school- en klaskenmerken op leerlingprestaties in taal en rekenen/wiskunde (bron: Scheerens en Bosker, 1997, p. 305; lege cellen wijzen op het ontbreken van overzichtsstudies ter zake)

	kwalitatieve syntheses	internationaal vergelijkend onderzoek	meta-analyses/ 'best-evidence' syntheses
'resource' factoren			
leerling-leerkracht ratio		-0.03	0.02
scholing leerkracht		0.00	-0.03
ervaring leerkracht			0.04
salaris leerkracht			-0.07
uitgaven per leerling			0.20
schoolorganisatiefactoren			
productief klimaat	+		
prestatiedruk m.b.t. 'basics'	+	0.02	0.14
onderwijskundig leiderschap	+	0.04	0.05
monitoring/evaluatie	+	0.00	0.15
samenwerking en consensus	+	-0.02	0.03
betrokkenheid ouders	+	0.08	0.13
stafontwikkeling	+		
hoge verwachtingen	+	0.20	
orderlijk klimaat	+	0.04	0.11
instructiefactoren			
gelegenheid om te leren	+	0.15	0.09
taakgerichte leertijd	+	0.00 (n.s.)	0.19
huiswerk		-0.01 (n.s.)	0.06
directe instructie	+	-0.01 (n.s.)	0.11 (n.s.)
coöperatief leren			0.27
feedback			0.48
reinforcement			0.58
differentiatie/adaptieve instructie			0.22

dat scholen elk jaar een goed inzicht krijgen voor welke vakken de prestaties achter blijven. Een tweede indicatie voor de maakbaarheid van de opbrengsten door de scholen kan ontleend worden aan het bepalen van die variabelen die variantie kunnen binden. Op basis van het onderzoek van Brandsma (1993) naar de rekenprestaties van leerlingen op de basisschool kan de variantie in de volgende componenten uiteengelegd worden (zie Tabel 6).

Tabel 6
Het binden van variantie in leerlingprestaties

	binnenschools	tussenschools
leerling covariabelen als ses, sekse, pretest, IQ, e.d.	50%	16%
schoolvariabelen	0%	10%
onverklaard	22%	2%

Van de oorspronkelijke 28% tussenschoolse variantie blijkt meer dan de helft te herleiden te zijn op verschillen in instroom van leerlingen met betrekking tot sociaal-economisch milieu van herkomst, sekse, intelligentie en parate

schoolgeschiktheid. Ongeveer eenderde deel van de tussenschoolse variantie kan herleid worden tot verschillen in de organisatie en het onderwijs van de scholen. Welke factoren dat zijn, is te zien in Tabel 7.

Te zien is een patroon waarbij factoren die verder weg liggen van het primaire proces ('resources' en schoolfactoren) een minder sterke samenhang met de opbrengsten te zien geven dan factoren die er rechtstreeks mee te maken hebben (instructiekenmerken). Interessant met het oog op schoolverbetering is de vraag of scholen zelf weten hoe ze hun opbrengsten kunnen verhogen. Uit het eerder genoemde onderzoek van Van Amelsvoort et al. (1993) blijken echter met name de scholen met de lagere opbrengsten meer geneigd hun toevlucht te zoeken in oneigenlijke maatregelen als verscherpen van het doublure-, doorstroom- en aannamebeleid. De scholen met de hogere opbrengsten daarentegen noemen vaak meer eigenlijke maatregelen als het optimaliseren van het functioneren van de schoolorganisatie, de leerlingbegeleiding, en het selectie- en determinatiebeleid alsmede vakinhoudelijke verbeteringen.

De onderwijscononoom Monk (1992), het door economen uitgevoerde onderzoek naar onderwijsproductiefuncties overziend (onderwijskundigen zouden dat onderwijseffectiviteitsonderzoek noemen), komt tot de slotsom dat het bewijs voor het bestaan van een dergelijke functie mager is. In dat opzicht lijkt het standpunt van de schoolleiders die oneigenlijke maatregelen voorstellen om de opbrengsten te verbeteren begrijpelijk. Maar Monk komt met een alternatief. Wellicht, zo stelt hij, bestaat er een quasi-productiefunctie, die maar één productiviteitsverhogende factor kent: de probleemoplossende capaciteit van de school. Scholen met een voldoende probleemoplossende capaciteit zijn in staat gegeven de eigen omstandigheden een optimale mix van maatregelen te treffen om de productiviteit te verhogen. Om de eigen opbrengsten naar waarde te kunnen schatten is het dan niet alleen verstandig dat scholen zich meten aan een benchmark (bijvoorbeeld het nationaal gemiddelde), maar dat zij zich ook kunnen oriënteren op benchmark scholen die in dezelfde omstandigheden verkeren.

2.4 Tussenbalans

In het voorafgaande is de betekenis van het openbaar maken van resultaten van scholen nogal gerelativeerd. Ouders gebruiken de gegevens die beschikbaar komen zelden als doorslaggevende informatie bij de schoolkeuze. Bovendien verschillen scholen zo weinig, dat er in veel gemeenten geen betekenisvolle verschillen in prestaties zullen zijn. Wat betreft schoolverbetering is er ook slechts een beperkte betekenis, namelijk vooral door een basis te creëren voor benchmarking (de school kan zich vergelijken met andere scholen en met het landelijk gemiddelde). Voor schooldiagnose en zelfevaluatie is meer gedetailleerde informatie nodig. Verder zijn de marges voor schoolverbetering smal en is de instrumentele kennisbasis ('wat werkt er?') nog onvolledig. Een wellicht niet bedoeld effect kan zijn dat scholen zich, vooral in situaties van harde concurrentie, door selectie gaan richten op verbetering van de gemiddelde prestaties. Het publiceren van resultaten van scholen in het kader van accountability (het ter verantwoording houden van scholen door een hoger bestuurlijk niveau) lijkt in Nederland tot nu toe geen duidelijk motief. Ervaringen uit de VS met deze toepassing wijzen uit dat de daadwerkelijke benutting tegenvalt en dat het vooral een symbolische functie heeft.

3 De Techniek

3.1 Technische onvolkomenheden

Van der Putten en Van der Kamp (1998) hebben reeds gewezen op onzuiverheden in de door Dronkers gebruikte gegevens en op enkele merkwaardigheden rond het toepassen van de factoranalyse die Dronkers uitvoert om een schaal voor het interne rendement te kunnen ontwikkelen. Het verweer van Dronkers dat de onvolkomenheden het gevolg zijn van fouten door de inspectie gemaakt, snijdt onzes inziens geen hout. Die gegevens zijn door de inspectie met een bepaald doel eind verzameld, namelijk om in het kader van haar taken een gesprek op gang te brengen tussen de school en de toezichthoudend inspecteur over in-, door-, uitstroompatronen en examenresultaten. Dat stelt andere eisen aan de gegevensverzameling dan wanneer dergelijke gegevens gebruikt zouden

worden voor evaluatieve doeleinden². In de tweede *Trouw*-publicatie gebruikt Dronkers zelfs uitstroomgegevens met de suggestie dat het hier om schoolverlaters gaat, hetzij harde drop out, hetzij leerlingen die om redenen die met het genoten onderwijs te maken hebben van school wisselen. Dergelijke vormen van uitstroom zijn in de inspectiegegevens echter niet te scheiden van schoolwisseling vanwege verhuizing. Met andere woorden: hier zijn ernstige validiteitsproblemen, zo ernstig dat het onverantwoord is deze uitstroomgegevens bij de rendementsberekeningen te betrekken. Validiteitsproblemen ontstaan voorts door het veronachtzamen van de bekende niveauverschillen in de eindexamens in het VBO en MAVO en de verschillen tussen de vakken in cijfering (zie ook: Luyten, 1998). Nergens houdt Dronkers bovendien rekening met de (aanzienlijke) sectorverschillen die zich in het VBO voordoen³. Een validiteitsprobleem ontstaat ook door de datagestuurde wijze waarop Dronkers zowel de eerste als de tweede keer zijn rendementsmaat bepaalt. Weliswaar komt een bijdrage door doublures aan de factorscore in de verkeerde richting in de tweede publicatie niet meer voor, maar wat blijft is natuurlijk dat de factoroplossing afhankelijk is van de in de gegevens aanwezige structuur, en dat die oplossing noch uniek noch in de tijd stabiel noch constant over de schoolsoorten heen is⁴.

3.2 Toegevoegde waarde

Het presenteren van gemiddelde examencijfers of doublurepercentages per school als indicatoren voor de opbrengst is een gevaarlijke onderneming. Men kan zich namelijk de vraag stellen: Representeert zo'n indicator de inspanning die een school geleverd heeft? Deze vraag heeft sinds de jaren zestig centraal gestaan in het onderzoek naar effectieve scholen. Daar is immers het probleem aan de orde of, bijvoorbeeld, schoolleiderschap invloed heeft op de prestaties van leerlingen. Om dit oorzakelijk verband te kunnen leggen, dienen we ons ervan te vergewissen dat hogere prestaties niet veroorzaakt worden door de gezinsachtergrond van de leerlingen in plaats van door schoolleiderschap. In dit kader werken onderzoekers op het gebied van schooleffectiviteit met het begrip toegevoegde waarde: wat voegt de

school toe aan het niveau dat de leerlingen bij instroom in de school reeds hebben?

Achtergrondkenmerken van leerlingen als etniciteit, sociaal-economisch milieu van herkomst en geslacht worden in het onderzoek naar effectieve scholen vaak als proxies (d.w.z.: benaderingen) gebruikt voor het startniveau wanneer directe metingen ter zake (zoals scores op een Cito-entreetoets) niet voorhanden zijn. In de praktijk van het publiceren van schoolresultaten, zowel in Nederland als in andere landen (bijvoorbeeld Australië, Frankrijk en sommige staten in de VS), wordt bij ontstentenis van gegevens over individuele leerlingen zelfs een groepsniveau-proxy gehanteerd. Als kenmerk van de leerlingpopulatie wordt bijvoorbeeld het aantal leerlingen dat in aanmerking komt voor 'free meals' genomen (of in Nederland het percentage leerlingen waarvoor de school faciliteiten heeft ontvangen in het kader van de culturele minderheden regeling). Vervolgens worden de opbrengsten die een school gerealiseerd heeft voor dit kenmerk gecorrigeerd. De vraag is echter of dit een adequate manier van werken is, en of we met de aldus gecorrigeerde opbrengstgegevens de toegevoegde waarde te pakken hebben. Veenstra et al. (1998) hebben dit onderwerp in hun kritiek op Dronkers ook reeds aan de orde gesteld, onder meer verwijzend naar de dissertatie van Bosker (1990), maar Dronkers snapt kennelijk het punt niet. Waar het om gaat is dat het werken met geaggregeerde gegevens "dangerous at best, and disastrous at worst" is (Aitkin & Longford, 1986, p. 42). Waar het om draait bij het bepalen van de door scholen toegevoegde waarde is de vaststelling van de gemiddelde mate waarin individuele leerlingen boven de statistische verwachting presteren op een school. Dat is iets geheel anders dan de vaststelling of een schoolpopulatie als geheel boven de statistische verwachting presteert.

Dit probleem heeft een conceptuele en een statistische achtergrond. Conceptueel kan het percentage leerlingen uit culturele minderheden voor iets anders staan dan een simpele sommatie van individuele omstandigheden. Het geeft voor leerlingen weer in welke schoolpopulatie ze zich bevinden, en als zodanig is het dus een kenmerk van de context van het onderwijs. Is het gemiddelde instroomniveau

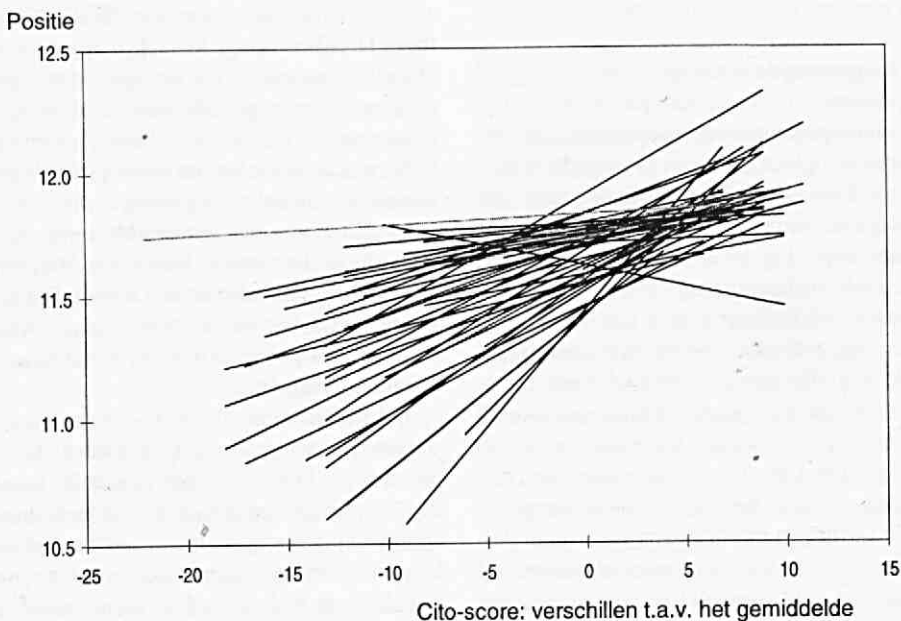
bijvoorbeeld aan de orde, dan kan dit geïnterpreteerd worden als een maat voor de rijkheid van de leeromgeving (Dar & Resh, 1986). Statistisch is er ook iets anders aan de orde als de relatie op leerlingniveau (relatie binnen scholen) tussen het instroomniveau en de examencijfers een andere sterkte en/of richting heeft dan de analoge relatie op het geaggregeerde niveau van de school (relatie tussen scholen). Correcties op leerling- dan wel op schoolniveau voor instroomverschillen vallen dan anders uit en scholen die bij de ene correctie positief uit de boot komen, komen dat dan in andere mate bij de andere correctie. In elk geval resulteert het uitvoeren van een correctie op geaggregeerd niveau zelden tot identieke resultaten als een correctie op leerlingniveau en in dat opzicht is het gebruik van de term toegevoegde waarde derhalve verwarrend.

3.3 Differentiële effecten

Tot nu toe hebben we het, zij het impliciet, alleen over de gemiddelde leerling gehad. Een school heeft hoge opbrengsten of juist niet. Maar binnen een school is er een enorme variatie: voor sommige leerlingen wordt wel en voor anderen juist niet een hoge opbrengst gerealiseerd. Dit is nu juist iets wat bij het schoolkeuzeproces een belangrijke rol zou moeten spe-

len: ouders vragen zich af of de school van hun voorkeur bij het kind past. Bijvoorbeeld: wat is de kans dat een leerling met een 'twijfel'-advies (HAVO-VWO) de betreffende school uiteindelijk met een VWO-diploma zal verlaten? De vraag is dus niet hoe goed de school het gemiddeld doet, maar wat de specifieke opbrengsten van de school voor dergelijke leerlingen zijn. Deze vraag verwijst naar wat we noemen differentiële schooleffecten: is een school gelijkelijk effectief voor de diverse te onderscheiden groepen leerlingen? De techniek om dit te onderzoeken ('slopes as outcomes') is wat lastig uit te leggen (de lezer zij verwezen naar Bosker & Snijders, 1990), maar we kunnen e.e.a. illustreren aan de gepresenteerde grafiek (zie Figuur 1).

De grafiek bevat voor elke school de lijn voor de regressie van de uiteindelijke schoolprestaties van een leerling op zijn of haar entreescore. Gemakshalve hebben we van elke entreescore het nationale gemiddelde afgetrokken, zodat de score 0 de score voor de gemiddelde leerling is. De grafiek maakt duidelijk dat een school die het goed doet voor de gemiddelde leerling het niet per se ook goed hoeft te doen voor leerlingen binnen die school die een van het gemiddelde afwijkende entreescore hadden. In dit voorbeeld zijn de verschillen tus-



Figuur 1. Differentiële opbrengsten van scholen

sen de scholen voor leerlingen die relatief laag op de entreetoets scoren het grootst. Voor een ouder is nu de vraag: gegeven de entreescore (bijvoorbeeld de score op de Cito-eindtoets), op welke school heeft mijn kind de beste vooruitzichten?

4 Toepassingscontexten en toegevoegde waarde

In het voorgaande hebben we stilgestaan bij de begrippen 'toegevoegde waarde' en 'differentiële toegevoegde waarde'. Het probleem dat Dronkers en ook de inspectie diende op te lossen was, dat er geen gegevens aanwezig waren over het instroomniveau van de leerlingen. In de eerder beschreven werkwijze hebben we aangegeven dat het eindniveau (examencijfers, behaald diploma, e.d.) te voorspellen is uit de parate en potentiële schoolgeschiktheid die de leerlingen hebben bij entree in de school voor voortgezet onderwijs. Het gaat dus om het cognitief niveau bij aanvang van de loopbaan in het secundair onderwijs. Dronkers gebruikt als proxy voor dat instroomniveau van een school het percentage leerlingen uit de culturele minderheden⁵. Daar is geen discussie over, zolang als er geen andere redenen zijn om deze variabele op te nemen. Dronkers echter verwijst naar het aan Bourdieu ontleende begrip 'cultureel kapitaal' (zie: De Graaf, 1987). Dat is echter een kenmerk van het gezin waarin het kind opgroeit, en daarmee lijkt het erop dat Dronkers deze variabele gebruikt niet alleen om te corrigeren voor het instroomniveau maar ook voor de onderwijsondersteuning die leerlingen in de thuissituatie genieten.

Dat is het essentiële probleem als er meer achtergrondgegevens van leerlingen bij de bepaling van de verwachte uitkomsten worden betrokken: een indicatie voor het sociaal-economisch milieu, bijvoorbeeld, naast een Cito-score kan alleen gelegitimeerd worden als men kan volhouden dat niet gemeten gezinsinvloeden (maar door het sociaal-economisch milieu van herkomst geïndiceerd) op de loopbaanontwikkeling van de leerling de meting van het werkelijke schooleffect verstoren. We raken hiermee aan een schier onoplosbaar probleem: de onmogelijkheid te komen tot een sluitende bewijsvoering over causale attributie, ofwel:

komt het door de school die ze bezoeken dat leerlingen het zo goed of slecht doen? Dit probleem is eerder opgemerkt door Raudenbush en Willms (1996) en door Grisay (1996). Het probleem ontstaat door keuzegedrag van leerlingen en hun ouders en door selectiegedrag van scholen (in plaats van dat leerlingen aselekt aan scholen worden toegewezen). Stel dat een school kwalitatief hoogwaardig onderwijs te bieden heeft (we doelen nu op het primaire proces en niet zozeer op de uitkomsten), en stel dat de leerlingen met een hoger schoolgeschiktheidsniveau en hun ouders bewust deze school kiezen⁶. Dan is het onderliggende model niet meer simpel input → proces → output, maar het wordt gecompliceerd tot proces → input → proces → output. Het statistisch model voor het berekenen van de toegevoegde waarde is echter gebaseerd op het eerste model, en de consequentie is, dat, als het tweede model geldig is, de correctie voor instroomverschillen tevens een onbedoelde correctie voor de kwaliteit van het onderwijsproces is. De implicatie is, dat men er verstandig aan doet zowel de bruto (bijv. ruwe examencijfers) als de netto opbrengstgegevens (de 'toegevoegde waarde', in casu examencijfers gecorrigeerd voor Cito-entreetoetscores) in de beschouwing te betrekken want daar ergens tussenin ligt de echte edoch onbekende bijdrage van de school⁷.

We komen nu bij een ander probleem, en dat is dat het begrip toegevoegde waarde en zijn operationalisering afhankelijk is van de toepassingscontext. In de context van 'accountability' en schoolverbetering is het streven erop gericht zo zuiver mogelijk de inspanning die de school gepleegd heeft boven tafel te krijgen. De correcties kunnen in dit geval niet streng genoeg zijn: alle gegevens (onveranderbare omstandigheden) dienen bij de correctie betrokken te worden, niet alleen de Cito-entreetoets, maar ook de gezinsachtergrond, de omgeving waarin de school staat, kenmerken van de schoolpopulatie als geheel e.d. Wat resteert is dan een goede benadering voor wat vermoedelijk door inspanningen van de kant van de school gerealiseerd is. Voor ouders is maar één ding interessant, en om dat duidelijk te maken keren we terug naar het onderscheid in binnen- en tussenschoolse regressie. We hebben betoogd dat de relatie tussen het eindniveau en

het startniveau de basis vormt voor het berekenen van toegevoegde waarde. Normaliter impliceert het de volgende regressievergelijking (we zien gemakshalve even af van differentieële schooleffecten):

$$\text{eind}_{is} = \text{intercept} + \beta_1 \text{ start}_{is} + R_{is} + U_s \quad (1)$$

Het subscript 1 staat voor leerling en s voor school. U_s is het schooleffect, ofwel de toegevoegde waarde. Het probleem zit hem nu in de regressiecoëfficiënt β_1 . Dit is een mixture van de binnen- en van de tussenschoolse regressie van het eindniveau op het startniveau. Die tussenschoolse regressie geeft het verband weer tussen het schoolgemiddelde eindniveau en het schoolgemiddeld startniveau en zoals we reeds hebben aangegeven kan men dit interpreteren als het effect van de leeromgeving. Maar ouders willen daar geen correctie voor: als hun kind goed leert omdat de schoolpopulatie gunstige karakteristieken heeft dan is dat mooi meegenomen.

Derhalve is het verstandiger dit effect expliciet te modelleren, zodat β_1 niet langer vervuild wordt door contextuele effecten.

$$\text{eind}_{is} = \text{intercept} + \beta_2 \text{ start}_{is} + \beta_3 \text{ start}_{is} + R_{is} + U_s \quad (2)$$

Als β_3 significant afwijkt van 0, dan is er inderdaad een contextueel effect. De toegevoegde

waarde waar ouders nu geïnteresseerd in zijn, bevat tevens het effect van de rijke leeromgeving:

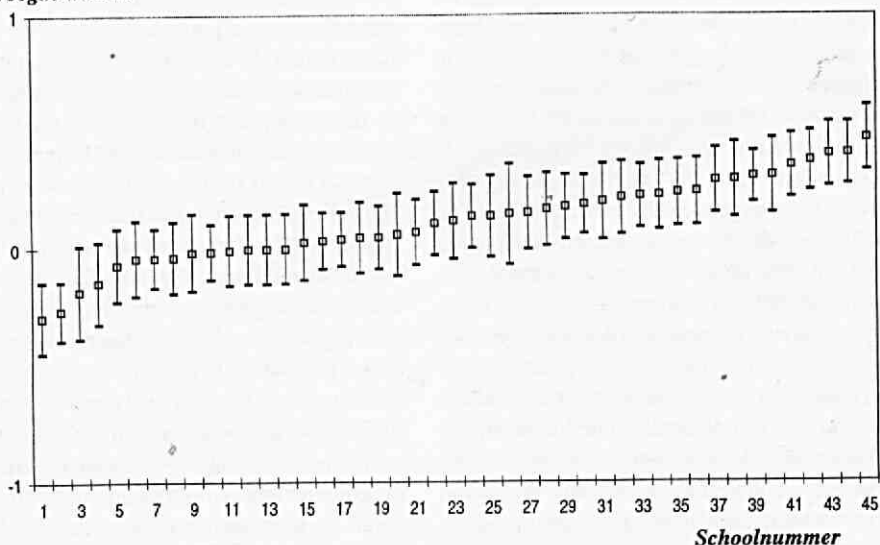
$$\beta_3 \text{ start}_{is} + U_s$$

In het algemeen is het overigens zo dat na correctie voor instroomverschillen de verschillen in netto opbrengsten van scholen nog eens aanmerkelijk kleiner zijn dan de toch al beperkte verschillen die we in het begin van deze bijdrage hebben laten zien. Om een indruk van de omvang van die verschillen in toegevoegde waarde te geven, is Figuur 2 opgenomen, waarin de prestaties van leerlingen in het MAVO na 3 jaar onderwijs zijn gemeten. Helaas hebben zittenblijvers de betreffende toetsen voor wiskunde en Nederlands niet gemaakt zodat dit een door 'attrition bias' aangetaste en dus niet representatieve onderzoeksgroep is⁸. Alleen scholen zijn opgenomen met minimaal 25 leerlingen.

We zien dat de overlap tussen de scholen groot is: er zijn maar weinig scholen die significant van elkaar verschillen. Bedenk daarbij dat in die gevallen waar het zich toch voordoet, het vaak voor een kiezende ouder geen reële alternatieven betreft omdat de scholen niet beide bereikbaar hoeven te zijn.

Er zitten overigens nog enkele adders onder het gras. Opbrengstgegevens gebaseerd op

Toegevoegde waarde



Figuur 2. Verschillen in toegevoegde waarde tussen scholen inclusief 90%-betrouwbaarheidsintervallen

onder andere eindexamenresultaten betreffen een lichte leerlingen die vier tot zes jaar geleden hun onderwijsloopbaan in het voortgezet onderwijs zijn begonnen. Het kind en zijn ouders die voor de schoolkeuze staan, hebben met de toegevoegde waarde dus een gegeven in handen waarvan ze maar moeten hopen dat dit ook voor de huidige instroomlichting gerealiseerd zal worden. Het is in dat opzicht verstandiger de toegevoegde waarde apart te berekenen voor de brugperiode en dan nog eens voor de bovenbouw. De tweede adder betreft de schoolwisselingen: als een redelijk deel van de eindexamenleerlingen op een bepaalde school hun eerste jaren in het voortgezet onderwijs op andere scholen heeft doorgebracht, welke school krijgt dan de credits?

5 Ten slotte

In deze bijdrage zijn de diverse toepassingscontexten voor het openbaar maken van schoolresultaten alsmede de technische kant van de berekening daarvan de revue gepasseerd. Geconstateerd kan worden dat de communis opinio is, dat het openbaar maken van schoolresultaten om uiteenlopende redenen verstandig is. In het publieke debat wordt meermalen gewezen op de mogelijke negatieve effecten, met name dan het strategische gedrag dat één en ander kan uitlokken van de zijde van de scholen. Smith (1995) heeft alle gevaren op een rij gezet. De voornaamste die in Nederland een rol spelen, en een mogelijke remedie, zijn:

1. Kortzichtige gerichtheid op gunstige scores voor de eerstvolgende publicatie (iedereen gaat over en er wordt tot in de treure geoefend op examens). Oplossing: vooruitzicht dat met de praktijk van het publiceren lang doorgedaan zal worden, en een redelijk omvangrijke reeks kengetallen in plaats van slechts één of twee.
2. De school richt zich alleen op de kengetallen en vergeet niet-kwantificeerbare relevante uitkomsten (zoals sociale vaardigheid, zelfstandig werken). Oplossing: produceren een kengetal voor extern rendement.
3. De school richt zich op selectie. Wanneer dat streng selecteren betekent, is toegevoegde waarde het kengetal dat dit probleem oplost. Wanneer scholen de systematiek van

toegevoegde waarde begrijpen, kunnen ze zich richten op de minst vaardige leerlingen qua startniveau: met deze groep is immers de grootste winst te boeken. Oplossing: vermeld ook steeds bruto scores op kengetallen.

4. Inperking van het innovatief vermogen. Dat is wellicht de grootste bedreiging, en alleen adequaat op te lossen door ervoor te zorgen dat de set kengetallen flexibel is en steeds een aantal nieuwe bevat.

In het algemeen zijn er voor de problemen die Smith aansnijdt dus wel oplossingen denkbaar. Belangrijk is dat men het publiceren van opbrengstgegevens niet als alleen zaligmakend ziet, en hiervoor is aangegeven dat binnen de diverse toepassingscontexten de nodige relativeringen gemaakt dienen te worden over de nuttigheid van deze praktijk.

Ondanks deze relativeringen moet het toch beschouwd worden als een nuttige praktijk die bijdraagt aan de rationaliteit van keuzegedrag, besturingsprocessen op scholen zelf en van het overheidsbeleid. Naar onze mening is vooral het feit dat de aandacht op prestatie-indicatoren wordt gelegd een belangrijke prikkel voor scholen om meer opbrengstgericht te gaan opereren. Een dergelijke wijziging in de schoolcultuur zal naar verwachting leiden tot prestatieverbetering.

Noten

1. Welk een revolutie zou er uitbreken als een leraar eens zo normloos zou beoordelen. Met de toegepaste methodiek – 'grading on the curve', met een curve die loopt van 1 tot en met 10 – heeft Dronkers schromelijk onderschat welk een maatschappelijke implicaties het uitdelen van rapportcijfers (want die begrijpt iedereen) aan scholen kan hebben. In dit opzicht is de tweede *Trouw*-publicatie een vooruitgang.
2. In die context is het niet bezwaarlijk dat de gegevens enigszins vervuild zijn. Men moet zich natuurlijk wel realiseren dat het hier om een kosten-batenafweging gaat: het construeren van een volledig betrouwbaar en valide gegevensbestand kost nu eenmaal menskracht.
3. Sommige vakken hebben om onduidelijke redenen een hoger gemiddelde voor het centraal

- eindexamen dan andere vakken. In 1997 deed zich dit bijvoorbeeld bij wiskunde A in het VWO voor, dat met 6.9 zeer hoog scoorde. Als een school nu 'toevallig' veel leerlingen heeft die wiskunde A volgen, dan wordt het schoolgemiddeld cijfer (Dronkers middelt over wiskunde A, Engels en Nederlands) kunstmatig hoog.
- 4 Het volledig bouwen op de resultaten van de factoranalyse leverde de eerste keer zelfs een resultaat op in het VWO waarin de richting van een bijdrage aan de rendementscore afhankelijk was van het leerjaar!
 - 5 In de kwaliteitskaarten van de inspectie is dat ook precies de reden voor de selectie van een aantal kenmerken (percentage leerlingen uit culturele minderheden, percentage leerlingen waarvan de ouders een tegemoetkoming in de studiekosten ontvangen, vier grote steden versus overig Nederland en denominatie) die tezamen als 'stand in' voor het onbekende instroomniveau fungeren (Inspectie van het onderwijs, 1998, p. 8).
 - 6 Mutatis mutandis ontstaat hetzelfde probleem als kwalitatief hoogwaardige scholen hun leerlingen gericht selecteren.
 - 7 Zulks onder de aanname dat niet één der betrokken partijen tegenovergestelde doelen nastreeft.
 - 8 De redenering van Van der Werf en Creemers dat dit juist goed is, is volstrekt onjuist en misleidend.

Literatuur

- Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A (General)* 149, (Part 1), 1-43.
- Amelvoort, H.W.C.H. van, Bosker, R.J., & Brandsma, T.F. (1993). Rendementsbekostiging in de bovenbouw van HAVO en VWO; overwegingen en reacties. *Pedagogische Studiën*, 70, 283-297.
- Bosker, R.J. (1990). *Extra kansen dankzij de school* (Academisch proefschrift). Nijmegen: ITS/OoMO.
- Bosker, R.J. (1998). Kwaliteit als toegevoegde waarde. In R.J. Bosker e.a., *Het vergelijken van scholen* (pp. 85-102). Enschede: OCTO/Universiteit Twente.
- Bosker, R.J., & Snijders, T.A.B. (1990). Statistische aspecten van multi-niveau onderzoek. *Tijdschrift voor Onderwijsresearch*, 15, 317-329.
- Brandsma, H.P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs*. Groningen: RION.
- Cibulka, J.G., & Derlin, R.L. (1995) State educational performance reporting policies in the U.S.: accountability's many faces. *International Journal of Educational Research*, 23, 479-492
- Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal*, 23, 357-374.
- Derriks, M., Kat E. de, & Deckers, P. (1995). *Schoolkeuzemotieven van kinderen bij de overgang bo/vo* (SCO-rapport 417). Amsterdam: SCO-Kohnstamm Instituut.
- Dronkers, J. (1998). Het betere is de vijand van het goede. Een reactie op de commentaren over het Trouw rapportcijfer. *Pedagogische Studiën*, 75, 142-150.
- Graaf, P.M. de (1987). *De inhoud van financiële en culturele hulpbronnen in onderwijsloopbanen*. Nijmegen: ITS/OoMO.
- Grisay, A. (1996). *Evolution des acquis cognitifs et socio-affectifs des élèves au cours des années de college*. Liège: Université de Liège.
- Inspectie van het Onderwijs (1998). *De kengtallen op de kwaliteitskaart. Een technische toelichting*. Leiden: SDU-servicecentrum.
- Lam, J.F. (1998). Publicatie van schoolgegevens: een internationale vergelijking. In R.J. Bosker e.a., *Het vergelijken van scholen* (pp. 11-33). Enschede: OCTO/ Universiteit Twente.
- Luyten, H. (1994). Stability of school effects in Dutch secondary education: The impact of variance across subjects and years. *International Journal of Educational Research*, 21, 197-216.
- Luyten, H. (1998). Moeilijke en minder moeilijke examens. In R.J. Bosker e.a., *Het vergelijken van scholen* (pp. 37-51). Enschede: OCTO/Universiteit Twente.
- Monk, D.H. (1992). *Education productivity research: An update and assessment of its role in education finance reform*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- Putten, C.M. van, & Kamp, L.J.T. van der, (1998). Dronkers' rapportcijfers voor de middelbare scholen: nauwkeurigheid, geldigheid en betekenis. *Pedagogische Studiën*, 75, 135-141.
- Raudenbush, S.W., & Willms, J.D. (1996). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20, 307-335.
- Scheerens, J. (1998). *Decentralization, school improvement and school evaluation*. Copenhagen: