



UvA-DARE (Digital Academic Repository)

Een economische kijk op publieke indicatoren van schoolkwaliteit

Oosterbeek, H.; Webbink, D.

Publication date
2001

[Link to publication](#)

Citation for published version (APA):

Oosterbeek, H., & Webbink, D. (2001). *Een economische kijk op publieke indicatoren van schoolkwaliteit*. Universiteit van Amsterdam.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

EEN ECONOMISCHE KIJK OP PUBLIEKE INDICATOREN VAN SCHOOLKWALITEIT¹

Hessel Oosterbeek en Dinand Webbink²

Februari 2001

1. Inleiding

Vijf jaar heeft het geduurd voordat de Britse *league tables* naar Nederland zijn overgewaaid. In 1992 werden de eerste ranglijsten van scholen in het Verenigd Koninkrijk gepubliceerd, in 1997 begon het dagblad Trouw met de publicatie van schoolprestaties van scholen voor voortgezet onderwijs in Nederland. Hoewel de nadelen en tekortkomingen van het openbaar maken van schoolprestaties in het VK al royaal onder de aandacht waren gebracht, bleek men daar in Nederland niet veel van geleerd te hebben. In het VK is de praktijk van de league tables onder anderen aan de kaak gesteld door Harvey Goldstein en coauteurs. Zij hebben zowel op methodologische, als op ethische en politieke aspecten gewezen.³ Onze indruk is eigenlijk dat grote delen van de Nederlandse discussie een herhaling zijn van de Britse discussie. In onze bijdrage willen we trachten enkele nieuwe elementen voor het voetlicht te brengen. Doordat beide auteurs van deze bijdrage van huis uit econoom zijn baseren we ons op recente inzichten uit de onderwijseconomische literatuur. Op hoofdlijnen gaat het ons om drie punten:

1. Ranglijsten suggereren causale bijdragen van scholen aan prestaties van leerlingen, deze suggestie wordt volstrekt niet waargemaakt. Recent onderzoek geeft interessante voorbeelden voor het vaststellen van causaliteit.
2. Individuele leerprestaties zijn een beperkte indicator. De 'output' van een school komt waarschijnlijk veel beter tot uitdrukking in de latere arbeidsmarktpositie van de leerlingen.

¹ Een iets gewijzigde versie van dit artikel is in 2001 gepubliceerd in het boek "Het oog der natie: scholen op rapport" onder redactie van A.B. Dijkstra, S. Karsten, R. Veenstra en A.J. Visscher, en uitgegeven door Van Gorcum in Assen.

² De eerste auteur is werkzaam bij de economische faculteit van de UvA, de tweede auteur is verbonden aan SCHOLAR (UvA) en de afdeling Kenniseconomie van het Centraal Planbureau.

³ Voor een overzicht zie Harvey Goldstein's homepage: www.ioe.ac.uk/hgpersonal/

3. Ranglijsten geven onbedoelde en bedoelde prikkels aan leerlingen, ouders, leerkrachten en schoolbesturen. Gezien de huidige kwaliteit van de indicatoren is bekostiging op output geen beleidsoptie.

Onze conclusie is dat de kwaliteit van de schoolindicatoren in de Trouw-ranglijsten onvoldoende is. Het terugdraaien van de huidige praktijk is echter geen optie. Met dat gegeven bespreken we in de slotparagraaf wat volgens ons vanaf nu het best kan worden gedaan.

2. Het meten van schoolprestaties

De prestaties van scholen worden in de onderwijseconomie geanalyseerd met onderwijsproductiefuncties. Daarin worden de prestaties van leerlingen gerelateerd aan ruwweg drie groepen factoren. Allereerst zijn dat de kenmerken van de leerling, zoals de initiële aanleg en de sociaal-economische achtergrond. In de tweede plaats zijn dat de kenmerken van de klas, zoals de docent, de groeps grootte en kenmerken van de medeleerlingen (peergroup). En, in de derde plaats zijn dat kenmerken van de school. Verder is in het model een storingsterm opgenomen die de invloed van niet-waargenomen factoren weergeeft die van invloed zijn op de prestaties van leerlingen.⁴ Dit kan gaan om niet-waargenomen kenmerken van leerlingen, ouders of school, maar ook om niet-waargenomen verschillen in voorkeuren alsmede om werkelijke toevalligheden in prestaties.

Veel van de kritiek op de ranglijsten, zoals door Trouw gepubliceerd, is direct af te leiden uit dit analysekader. In de ranglijsten wordt onvoldoende rekening gehouden met de andere, dan de schoolse, factoren. In de ranglijsten wordt gewerkt met het eindniveau van de leerlingen zonder rekening te houden met verschillen in startniveau. Er wordt geen aandacht besteed aan de statistisch onzekerheid rond de prestaties van scholen. Omdat deze kritiek ook in diverse andere hoofdstukken aan de orde is (...) en tevens door Goldstein en anderen is verwoord gaan we hier niet verder op in.

Eigenlijk gaat het om het vaststellen van causale effecten

Een belangrijke recente ontwikkeling in de onderwijseconomie is het inzicht dat veel klas- en schoolkenmerken, die van invloed zijn op de prestaties van leerlingen, niet toevallig over leerlingen zijn gespreid maar systematisch samenhangen met niet-waargenomen kenmerken van de leerlingen die op hun beurt weer systematisch samenhangen met (verwachte) leerprestaties. Dit inzicht is van groot belang als onderzoekers trachten de causale effecten van kenmerken van de school of de klas

⁴ Bij het schatten van deze modellen wordt er rekening mee gehouden dat de niet waargenomen factoren voor leerlingen in dezelfde klas of op dezelfde school gecorreleerd zijn.

op de prestaties van leerlingen te bepalen. Als hiermee in de analyse geen rekening wordt gehouden kunnen veranderingen in leerprestaties ten onrechte worden toegeschreven aan schoolkenmerken terwijl ze in werkelijkheid veroorzaakt worden door andere factoren. Een voorbeeld is de invloed van de groepsgrootte op prestaties van leerlingen. Stel dat ouders die veel aandacht besteden aan de leerprestaties van hun kinderen een voorkeur hebben voor scholen met kleine klassen en dat deze ouderlijke betrokkenheid ook een onafhankelijk effect heeft op de prestaties van de kinderen. In dat geval bestaat de kans dat kinderen met sterk betrokken ouders vaker in relatief kleine klassen zitten. De onderzoeker neemt echter de grotere betrokkenheid van ouders niet waar. In het traditionele analysemodel zou het effect op de leerprestaties worden toegeschreven aan de groepsgrootte terwijl dit in werkelijkheid veroorzaakt wordt door de (niet waargenomen) betrokkenheid van de ouders. De groepsgrootte wordt in dit geval een endogene variabele genoemd. Het geschatte effect van de groepsgrootte op de leerlingprestaties wordt vertekend als de groepsgrootte samenhangt met een niet waargenomen variabele (de betrokkenheid van de ouders) die tevens van invloed is op de prestaties van de leerling. Voor vrijwel alle klas- en schoolkenmerken geldt dat ze samen kunnen hangen met niet-waargenomen kenmerken die op hun beurt direct van invloed zijn op uitkomsten. Vertekening van het causale effect van deze klas- en schoolkenmerken op uitkomsten ligt dan voortdurend op de loer. Wie dit gevaar veronachtzaamt kan in specifieke gevallen gelijk hebben, het probleem is echter dat dit pas kan worden vastgesteld als is onderzocht of de vertekening optreedt.

Dit inzicht is uiteraard van groot belang bij de beoordeling van de ranglijsten, daar gaat het immers om de causale bijdrage van scholen aan leerlingprestaties. Als school A hoger scoort dan school B wil dat niet meer zeggen dan dat gemiddeld genomen - en gecorrigeerd voor enkele achtergrondkenmerken - de leerlingen die voor school A hebben gekozen het beter doen dan de leerlingen die voor school B hebben gekozen. Dat is heel iets anders dan dat voor een willekeurige leerling geldt dat deze het beter zou doen op school A dan op school B. Op deze manier bekeken hebben we hier te maken met een standaard evaluatie-probleem. We willen weten of de ene treatment (school A) beter werkt dan de andere treatment (school B). Als er geen random toewijzing van leerlingen over scholen A en B plaatsvindt, en als er ook geen andere bronnen van exogene variatie zijn die de keuze tussen scholen A en B beïnvloeden, is het eigenlijk niet goed mogelijk om het verschil in effect tussen treatments te meten. We zijn dan aangewezen op andere methoden die erop neerkomen dat zoveel mogelijk andere covariaten constant worden gehouden. "Matching" is een van die methoden, en multivariate analyse een andere. En de correctie voor enkele achtergrondkenmerken zoals in de ranglijsten gebeurt, is daar ook een poging toe. Inmiddels zijn er nogal wat studies die

laten zien dat deze manieren van correctie tot erg misleidende conclusies kunnen leiden⁵. Het probleem is echter nog groter. Zoals eerder genoemd is de samenstelling van de groep van medeleerlingen een belangrijke omgevingsfactor die van invloed is op de leerprestaties van een kind. Deze factor zou ook constant moeten worden gehouden. De uitspraak dat school A beter is dan school B claimt dan eigenlijk dat als de leerlingpopulaties van beide scholen worden omgewisseld, de kinderen die van B naar A zijn gegaan allemaal beter gaan presteren, en dat de kinderen die van A naar B zijn gegaan allemaal minder gaan presteren. Dit kunnen we nooit nagaan. En als dat niet voor alle leerlingen geldt, zullen de winsten en verliezen van verschillende leerlingen op een of andere manier met elkaar vergeleken moeten worden.

Het bovenstaande lijkt te suggereren dat de methodologische problemen zo groot zijn dat ze welhaast onoplosbaar zijn. En dat zou dan betekenen dat vergelijking van scholen onmogelijk is. Zo erg is het gelukkig niet. De ideale methode om de prestaties van scholen met elkaar te vergelijken is opzet van een veldexperiment waarbij leerlingen door het lot aan scholen worden toegewezen. Dat ideaal is waarschijnlijk niet realiseerbaar. Toch is het goed om langs deze lijn over het probleem na te denken en op zoek te gaan naar omstandigheden die zo'n experiment zo natuurgetrouw mogelijk nabootsen. Zulke omstandigheden worden in de recente economische literatuur "natuurlijke experimenten" genoemd. Een fraai en overtuigend voorbeeld wordt beschreven door Angrist en Lavy (1999). Zij gebruiken een discontinuïteit in het bekostigingssysteem van Israëliëse basisscholen om het causale effect van klassenverkleining op leerprestaties te schatten.⁶ De discontinuïteit die zij benutten is een regel van de 12de eeuwse rabbijn Maimonides die zegt dat er tot 40 leerlingen in een klas kunnen zitten, maar dat bij het 41ste kind er een leerkracht bij moet komen. Deze regel speelt nu nog een belangrijke rol in de onderwijsbekostiging in Israël. Hierdoor bestaan duidelijke discontinuïteiten in de klassengrootte als functie van het aantal leerlingen per leerjaar. Bestaat een cohort uit 40 leerlingen dan is er één klas van 40 leerlingen. Bestaat het cohort uit 41 leerlingen dan zijn er twee klassen van gemiddeld 20½ leerlingen. Deze plotselinge overgang kan benut worden om het effect van een verkleining van klassen te schatten onder de aanname dat ouders niet weten of het voor hun kind relevante cohort uit 40 of 41 leerlingen zal bestaan.

Voor het vergelijken van scholen kan een vergelijkbare discontinuïteit worden benut. Ironisch genoeg hebben we hier het oog op discontinuïteiten die zijn veroorzaakt door de publicatie van ranglijsten van scholen. Voor de toelating tot een bepaald schooltype van voortgezet onderwijs

⁵ zie bijvoorbeeld LaLonde (1986).

dient een leerling op de cito-toets minstens een bij dat schooltype behorende drempelwaarde te hebben gescoord. Het geval wil nu dat sommige scholen deze toelatingsscores nog enigszins hebben verhoogd. Kinderen die eigenlijk naar die school wilden maar te laag hebben gescoord worden nu niet toegelaten en moeten naar een andere school. Deze leerlingen zijn als het ware uitgeloot. Door de prestaties van deze leerlingen op de school van hun tweede keuze te vergelijken met de prestaties van leerlingen die wel zijn toegelaten, kunnen verschillen tussen scholen worden gemeten. In de analyse dient verder wel gewoon voor de cito-score als indicator van de initiële aanleg gecontroleerd te worden. Zonder dat te doen zou de afwijzende school wel erg in het voordeel zijn. De identificerende aanname is dan dat er zich in de relatie tussen prestatie en cito-score geen discontinuïteiten voordoen bij het punt van de drempelwaarde. (Tevens moet worden aangenomen dat de afwijzing geen effect heeft op de motivatie van de leerlingen.) Dit is één voorbeeld van een manier op basis waarvan werkelijke verschillen tussen scholen kunnen worden gemeten. Wij vermoeden dat met een combinatie van creativiteit en voldoende gedetailleerde kennis van het onderwijsveld er nog (veel) meer van dit soort natuurlijke experimenten zijn te evalueren. Wie echt geïnteresseerd is in verschillen tussen scholen zou naar ons idee langs deze weg verder moeten.

De positie op de arbeidsmarkt als indicator

Een ander belangrijk kritiekpunt op de door Trouw gekozen aanpak van ranglijsten is dat de gekozen prestatie-maten geen recht doen aan het multidimensionele karakter van schoolprestaties.

De bijdragen van scholen omvatten veel meer dan hetgeen gemeten wordt met de indicatoren van Trouw. Uitbreiding van indicatoren ligt dan voor de hand. In de visie van economen zou deze uitbreiding zich niet moeten beperken tot testcores of vaardigheidsmaten bij het verlaten van het onderwijs maar vooral informatie moeten opleveren over het latere functioneren van de leerling in de samenleving. Het leren op school is immers allereerst bedoeld voor dit latere functioneren in de samenleving. Een indicator waarin het multidimensionele karakter van de 'output' van een school tot uitdrukking komt en die tevens informatie verschaft over het functioneren in de samenleving is *de arbeidsmarktpositie*, bijvoorbeeld het loon of de kans op een baan van de leerling. Toegegeven, ook deze indicator is niet perfect omdat het allerhande opbrengsten van onderwijs in de privé-sfeer buiten beschouwing laat. Maar toch zijn er veel voordelen: vele opgedane vaardigheden worden meegewogen (buiten de cognitieve vaardigheden ook sociale en emotionele vaardigheden), de gewichten die aan de verschillende vaardigheden worden toegekend zijn niet willekeurig tot stand

⁶ In zekere zin is deze toepassing van het idee van natuurlijke experimenten tevens een toepassing van de in de sociale wetenschappen al langer bekende methode van de "regression discontinuity design" (zie Campbell 1969).

gekomen maar kunnen gezien worden als de prijzen die op de arbeidsmarkt impliciet voor deze vaardigheden worden betaald. En nog een belangrijk voordeel: als scholen worden afgerekend op een beperkt aantal meetbare maatstaven, is niet uitgesloten dat er "teaching to the test" zal plaatsvinden. En zolang de test niet alles bevat dat relevant is, kan dat er toe leiden dat aan sommige onderdelen teveel en aan andere onderdelen te weinig aandacht wordt gegeven.

In diverse recente studies komt bovendien naar voren dat het in beschouwing nemen van de indicator 'arbeidsmarktpositie' een ander beeld geeft dan de indicator 'leerwinst'. Veel onderzoek naar de effecten van voor- of vroegschoolse interventies gaf teleurstellende resultaten. De aanvankelijke winst, in termen hogere IQ-scores, doofde na verloop van tijd weer uit. Als echter naar lange termijn effecten gekeken wordt, worden in verschillende studies wel degelijk gunstige effecten gevonden in termen van een hoger loon, een grotere kans op werk en minder kans op inactiviteit⁷. Een ander voorbeeld is het onderzoek naar de effecten van hogere onderwijsuitgaven. Lange tijd bestond het beeld dat verhoging van de onderwijsuitgaven, bijvoorbeeld door kleinere klassen, hogere salarissen of meer opleiding voor leraren, nauwelijks leidde tot verbetering van de leerprestaties van leerlingen. In diverse studies werden teleurstellende effecten gevonden van hogere onderwijsuitgaven op de prestaties van leerlingen (zie bijvoorbeeld Hanushek 1996). Met andere woorden, investeringen in onderwijskwaliteit leveren, in termen van leerwinst, weinig op. Dit beeld verandert wanneer gekeken wordt naar het effect op de arbeidsmarktpositie. Card en Krueger (1996) vinden substantiële positieve effecten van verbetering van de schoolkwaliteit (door hogere onderwijsuitgaven) op de latere arbeidsmarktpositie. Dit duidt erop dat de indicator 'leerwinst' te beperkt is. In de indicator arbeidsmarktpositie komen ook minder goed meetbare zaken als sociale en emotionele vaardigheden tot uitdrukking.

Tegenover het voordeel van de meer adequate meting van de bijdragen van scholen staat het nadeel dat informatie over de arbeidsmarktprestaties pas jaren later beschikbaar komt. Dat is de prijs voor de betere kwaliteit van de informatie. Naar onze mening is aan deze vertraging niet veel te doen. Na verloop van tijd kan wel duidelijk worden in hoeverre bepaalde scholen een bepaalde traditie opbouwen ten aanzien van het goed terecht komen van hun leerlingen. In hoeverre deze traditie een aanwijzing is voor de actuele kwaliteit is echter moeilijk te zeggen. Dit illustreert vooral dat informatie gemeten bij het verlaten van het onderwijs beperkt is en dat het blindvaren op indicatoren zoals de Trouw-ranglijsten gevaarlijk is.

3. Effecten van de ranglijsten

⁷ Zie Currie (2001), voor een recent overzicht.

De publicatie van de ranglijsten van schoolprestatie-indicatoren geven prikkels aan ouders/leerlingen, leerkrachten en scholen. Het is heel goed mogelijk dat de huidige misleidende informatie over verschillen in schoolkwaliteit op termijn zal leiden tot 'echte' verschillen in schoolkwaliteit. De informatie kan namelijk bepaalde ouders/leerlingen en leerkrachten stimuleren om naar de goed scorende scholen te gaan.

Ouders en leerlingen zullen de gegevens laten meewegen bij hun schoolkeuze, en lijken dit volgens onderzoek van Trouw ook werkelijk te doen⁸. Ze worden daardoor geprikkeld om hun kinderen naar een school te sturen die een goed scoort. Al is die goede score onterecht (en er zijn tal van aanwijzingen/aneddotes dat de eerste ranglijsten veel fouten bevatten), dan nog kan het ertoe leiden dat op den duur de school die een goede score ontving het beter doet dan andere scholen. Er doet zich dan een self-fulfilling prophecy voor omdat de school met een goede score de kinderen aantrekt van ouders die de schoolkeuze belangrijk vinden (belangrijk genoeg althans om de ranglijsten te bekijken). Als deze ouders ook meer energie steken in de begeleiding van hun kinderen (en onwaarschijnlijk klinkt dat niet), dan krijgt deze school vanzelf de betere leerlingen. En alleen maar corrigeren voor het percentage allochtonen en het percentage leerlingen dat een tegemoetkoming in de studiekosten krijgt, is onvoldoende.

Wat zijn de prikkels voor leerkrachten? Het lijkt ons plausibel dat een leerkracht liever werkt bij een school die goed scoort dan bij een school die slecht scoort. Scholen voor voortgezet onderwijs zijn tamelijk groot, dus is de gelegenheid voor individuele leerkrachten om de prestaties van de school te beïnvloeden gering. Om toch bij een school terecht te komen die goed te boek staat, zal de leerkracht die bij een "slechte" school werkzaam is dus van school moeten veranderen. Kortom, het zal scholen die als goed te boek staan minder moeite kosten om personeel te werven dan scholen die van Trouw - al dan niet terecht - een slechte beoordeling ontvingen. Dit versterkt verder de self-fulfilling prophecy.

Het idee is om scholen met het beoordelen en publiceren van hun kwaliteit een prikkel (incentive) te geven om beter te functioneren. Het uitdelen van incentives is een subtiele aangelegenheid, waaraan in de economische theorie veel aandacht is besteed in het kader van de zogenoemde principal agent theorie. In die theorie draait het om twee actoren: een principaal en een agent. De principaal doet een beroep op de agent om een bepaalde taak uit te voeren. (Het ministerie van onderwijs vraagt scholen om leerlingen te onderwijzen.) Complicaties doen zich voor doordat de principaal en de agent niet

⁸ zie Trouw, 11 mei 1999.

(precies) dezelfde doelstellingen hebben, en doordat de agent over informatie beschikt waar de principaal geen beschikking over heeft. (De minister van onderwijs legt nadruk op het onderwijzen van taal en rekenen terwijl de directeur van de school de sociale en emotionele ontwikkeling belangrijker vindt. De minister kan niet exact waarnemen hoeveel inspanning de leraren van de school leveren om de leerlingen tot het door de minister gewenste taal- en rekenniveau op te leiden.) De principaal dient er tevens rekening mee te houden dat de agent risico-avers is. (Als de minister de school aan teveel risico blootstelt, wil niemand - op een enkele goudzoeker na - het onderwijs in.) De principaal ziet zich nu voor de vraag gesteld in hoeverre zij de beloning van de agent zal laten afhangen van een nog te bepalen prestatie maatstaf. Het ene uiterste is dat de beloning helemaal niet afhangt van de gemeten prestaties; er is een vast salaris. Het andere uiterste is dat de beloning volledig afhangt van de gemeten prestaties; een voorbeeld is stukloon. Uit de literatuur komt naar voren dat het aandeel van de beloning dat afhangt van de prestaties zou moeten variëren met (i) de mate waarin de inspanning van invloed is op de gemeten prestaties, (ii) de nauwkeurigheid waarmee de inspanning kan worden afgeleid uit de gemeten prestaties, (iii) de risico-geneigdheid van de agent, en (iv) de gevoeligheid van de agent voor prikkels.⁹

Deze resultaten kunnen vertaald worden naar de relatie tussen de minister van onderwijs en scholen, en de invloed van ranglijsten daarop. Veel van wat we eerder bespraken blijkt daarbij inpasbaar. We gaan er hierbij voor de eenvoud vanuit dat een school als eenheid beschouwd kan worden, en zien daarmee af van principaal-agent problemen binnen scholen. Laten we eerst eens nagaan hoe de factoren (i) tot en met (iv) zich vertalen in de context van het onderwijs. De eerste factor heeft betrekking op de invloed van de mate waarin leerkrachten zich inspannen op de prestaties van leerlingen. Een van de oorzaken van het feit dat deze invloed niet al te groot is, is het feit dat leerlingen (en ouders) hun eigen inspanning aanpassen aan de inspanning van de leerkracht. Hoe meer de leerkracht zich inspannt des te minder zullen leerlingen en hun ouders dat doen.¹⁰ Wat de tweede factor aangaat is bekend dat van de totale variantie in leerlingprestaties hooguit 15-20% wordt veroorzaakt door de school. De rest komt van "buiten" in de vorm van zaken waar de school geen invloed op heeft (kenmerken van de leerlingpopulatie en toevallige factoren). Wat betreft de risicohouding is de aanname meestal dat werknemers risico-avers zijn; er zijn aanwijzingen dat ambtenaren (waaronder leraren) meer risico-avers zijn dan anderen. Wat de vierde factor - tot slot -

⁹ We kunnen dit uitdrukken in een elegante formule (vergelijk Milgrom en Roberts 1992, p.221): $b = P'(e) / [1 + rVC''(e)]$, hierin is b het aandeel van de beloning dat afhangt van de prestaties, $P'(e)$ is de invloed van inspanning (e) op de prestatie (P), r is een maat voor de agent's risico-houding (hoe hoger r hoe meer risico-avers de agent is), V is de variantie van de prestatie bij gegeven inspanningsniveau en C'' is de afgeleide van de marginale kosten van inspanning voor de agent.

betreft, bestaan er aanwijzingen dat werknemers sterk op prikkels reageren. Wellicht is deze gevoeligheid iets geringer onder leraren dan onder andere werknemers - ze hebben immers gekozen voor een beroep waarin de beloning niet erg hoog is.

Tot nu toe gingen we er in deze paragraaf vanuit dat de door de minister gewenste prestaties van scholen goed meetbaar zijn. Maar zoals we al eerder bespraken is dat niet zo. Sommige aspecten die wel van belang worden geacht vallen buiten de gemeten prestaties. Als scholen wel via het keuzegedrag van ouders en leerlingen worden afgerekend op deze gemeten prestaties, zullen ze zich minder inspannen voor de curriculumdelen die daarbuiten vallen. Er vindt dan "teaching to the test" plaats. Of er is sprake van - om het in de woorden van Kerr uit te drukken "[On] the folly of rewarding A, while hoping for B" (1975). Zo gaat het in de huidige ranglijsten vooral om het snel (onvertraagd) doorlopen van het onderwijstraject. Dit kan scholen stimuleren om zich vooral in te spannen voor leerlingen 'in de gevarezone' en minder aandacht te besteden aan de zwakste of sterkste leerlingen. Ook is het mogelijk dat scholen leerlingen aanzetten om minder risicovolle trajecten/ vakkenpakketten te kiezen (meer 'feestpakketten'). Een andere mogelijkheid is dat scholen op één of andere wijze leerlingen gaan selecteren. Dit kan er toe leiden dat 'problematische' leerlingen moeilijk een school kunnen vinden. Of nog erger, scholen kunnen aangezet worden tot fraude ten aanzien van de prestaties van hun leerlingen.

Al met al leidt deze beschouwing wat ons betreft tot de conclusie dat ten aanzien van scholen er maar een beperkte rol is weggelegd voor bekostiging op basis van gemeten prestaties. Kennelijk is dit ook de mening van de wetgever. Hierop wijst althans het feit dat de bekostiging van scholen in het basisonderwijs en in het voortgezet onderwijs uitsluitend afhangt van het aantal leerlingen en niet van de gemeten prestaties. Natuurlijk hangt het aantal leerlingen wel af van de onderwijskwaliteit die een school in de perceptie van ouders levert. In deze perceptie worden vele zaken op een of andere wijze meegewogen. Door de publicatie van ranglijsten wordt deze weging - naar wij vermoeden - verschoven in de richting van gemeten prestaties. Daarmee wordt de rol van prestaties in de bekostiging waarschijnlijk groter dan gewenst zou zijn op basis van de overwegingen die we hebben geschetst.

4. Wat dan wel?

¹⁰ Zo zullen ouders hun kinderen meer bij hun huiswerk helpen als ze de indruk hebben dat de leerkracht onvoldoende succes boekt.

Op zich is de wens naar informatie over de kwaliteit van scholen gerechtvaardigd en begrijpelijk. De overheid geeft scholen geld en wil weten of dit geld goed wordt besteed (accountability). Ouders en scholieren hebben behoefte aan informatie ten behoeve van de schoolkeuze. Beide partijen zijn echter niet geholpen met het soort ranglijsten dat nu jaarlijks in twee landelijke dagbladen verschijnt. Tegelijkertijd lijkt de huidige praktijk van het publiceren van ranglijsten van schoolprestaties onomkeerbaar te zijn. De vraag is dan hoe hier verstandig mee om te gaan. Naar onze mening dienen de gegevens op veel punten te worden verbeterd en dient ook veel beter uitgelegd te worden wat de betekenis is van deze gegevens (en wat er niet geconcludeerd mag worden).

Noodzakelijke verbeteringen

Een *eerste* noodzakelijke verbetering van de indicatoren van schoolprestaties is dat het moet gaan om 'value-added'-gegevens. Niet het leerniveau maar de leerwinst dient gemeten te worden. In de *tweede* plaats dienen de gegevens het multidimensionele karakter van schoolprestaties tot uitdrukking te brengen. Dit betekent dat gegevens gepresenteerd moeten worden over een brede range van vakken waarbij niet alleen de cognitieve aspecten worden gemeten maar ook de niet cognitieve aspecten van het leren (gedrags- en motivatie-aspecten). Daarbij zou tevens de prestaties voor bepaalde groepen leerlingen vermeld kunnen worden, bijvoorbeeld de leerwinst van leerlingen uit minderheden versus de autochtone leerlingen. Voorts dient daarbij rekening gehouden te worden met verschillen in de sociale achtergrond van leerlingen alsmede verschillen in hun voor- en buitenschoolse activiteiten. Een *derde* noodzakelijke verbetering is dat betrouwbaarheidsintervallen gepresenteerd worden bij de prestaties van scholen. Alleen dan kunnen prestaties van scholen worden vergeleken. Een *vierde* noodzakelijke verbetering is dat de mobiliteit van leerlingen tussen scholen in kaart moet worden gebracht. Hoe zit het precies met de in- en uitstroom van leerlingen in de eerste onderwijsjaren. De gegevens over de prestaties van scholen zouden sterk verbeteren als ook de latere ervaringen van leerlingen gepresenteerd zouden worden, bijvoorbeeld de prestaties in het vervolgonderwijs of, en nog veel beter, de ervaringen op arbeidsmarkt van leerlingen¹¹.

Tot slot, naar onze mening is het cruciaal om bij de presentatie van de gegevens nadrukkelijk te melden dat we, bij de huidige stand van kennis, niet precies kunnen aangeven hoe groot de bijdrage van de school is aan de scores op de verschillende indicatoren.

Referenties

¹¹ In de Verenigde Staten wordt dit type informatie door instellingen in het hoger onderwijs regelmatig via hun WEBSITE aangeboden. Vermeld worden onder meer de mediaan van het salaris van studenten die een baan geaccepteerd hebben alsmede het deel van de studenten dat een promotietraject succesvol afrondt (zie Romer 2000).

- Agerbeek, M. (1999), Rendement speelt nu rol bij schoolkeuze, *Trouw*, 11 mei, 1999.
- Angrist J. en V. Lavy (1999), Using Maimonides Rule to Estimate the effect of class size on children's academic achievement, *Quarterly Journal of Economics* **114**, 533-575
- Campbell, D.T. (1969), Reforms as experiments, *American Psychologist* **24**, 409-429.
- Card, D.& A.B. Krueger (1996), Labor market effects of school quality: theory and evidence, NBER working paper 5450.
- Currie, J. (2000), Early Childhood intervention programs: what do we know?, te verschijnen in *Journal of Economic Perspectives*.
- Goldstein, H. (1997), Value added tables: the less-than-holy grail, *Managing Schools Today* **6**, 18-19.
- Goldstein, H. & D.J. Spiegelhalter, League tables and their limitations: statistical issues in comparisons of institutional performance, zie www.ioe.ac.uk/hgpersonal
- Goldstein, H. & S. Thomas (1996), Using examination results as indicators of school and college performance, *Journal of the Royal Statistical Society A* **159**, 149-163.
- Hanushek, E.A. (1996), School resources and student performance, in: Burtless, G. ed. (1996), Does money matter? The effect of school resources on student achievement and adult success, Brookings Institution Press, Washington, D.C.
- Hoxby, C. (2000), Peer effects in the classroom: learning from gender and race variation, NBER working paper 7867.
- Kerr, S. (1975), On the folly of rewarding A, while hoping for B, *Academy of Management Journal* **18**, 769-783.
- LaLonde, R. (1986), Evaluating the econometric evaluations of training programs with experimental data, *American Economic Review* **76**, 604-620.
- Lazear, E.P. (1999), Educational production, Stanford University, september 1999, unpublished paper
- Milgrom, P. en J. Roberts (1992), *Economics, Organization and Management*, New Jersey: Prentice Hall.
- Romer, P.M. (2000), Should the government subsidize supply or demand in the market for scientists and engineers?, NBER working paper 7723.