

## Hoofdstuk 2

### Schoolprestatie-indicatoren in het voortgezet onderwijs: start, reacties en vervolg

Jaap Dronkers & René Veenstra

Graag citeren als: Dronkers, J., & Veenstra, R. (2001). 'Schoolprestatie-indicatoren in het voortgezet onderwijs: start, reacties en vervolg'. In A.B. Dijkstra, S. Karsten, R. Veenstra, & A.J. Visscher (red.). *Het oog der natie. Scholen op rapport. Standaarden voor de publicatie van schoolprestaties*. (pp. 21-35). Assen: Van Gorcum.

#### 1. Inleiding

Hoewel publieke verantwoording van de kwaliteit van scholen voor Nederland geen volstrekt nieuw onderwerp was, werd het door een publicatie van het dagblad Trouw in 1997 voor iedereen zichtbaar op de kaart gezet. Het debat dat sindsdien over de voors en tegens van publieke informatie over onderwijskwaliteit wordt gevoerd draagt er op veel punten nog steeds de sporen van. Om die reden beschrijft dit hoofdstuk de prestatiegegevens die Trouw publiceerde en de voornaamste reacties die daarop volgden (paragraaf 3). In paragraaf 2 wordt eerst een kort overzicht van de voorgeschiedenis gegeven, en paragraaf 4 zet enkele ontwikkelingen nadien op een rij. Paragraaf 5 biedt bovendien een eerste afweging van de plussen en minnen van deze eerste publicatie van prestatiegegevens van scholen.

#### 2. De publicatie van inspectiegegevens door Trouw

##### 2.1 De voorgeschiedenis<sup>1</sup>

In juni 1996 begon de Trouw-journaliste Agerbeek met een vooronderzoek naar de mogelijkheden om prestaties van scholen openbaar te maken. Zij begon met een ronde langs de Informatiseringsbank, de Inspectie van het Onderwijs en het ministerie van Onderwijs. In juli van datzelfde jaar deed zij een beroep op de Wet Openbaarheid van Bestuur (WOB) om de nodige gegevens te verkrijgen. Dit was nodig omdat het ministerie weigerde gegevens over schoolprestaties aan Trouw te verstrekken. De weigering was gebaseerd op twee gronden. Het doorgeven van het materiaal aan derden zou allereerst de vertrouwensrelatie van de Inspectie met de scholen aantasten. Ten tweede zouden de scholen schade ondervinden van publicatie van de gegevens, omdat slagingspercentages geen rekening houden met bijvoorbeeld het beginniveau of de sociaal-economische achtergrond van leerlingen. In augustus 1997 stelde de rechter Agerbeek in het gelijk. De vertrouwensrelatie van de inspectie met de scholen kon volgens de rechter niet in gevaar zijn, omdat scholen wettelijk verplicht zijn de inspectie gegevens te leveren. Verder vond de rechter het publieke belang van het vrijgeven van cijfers zwaarder wegen dan de bedenkingen die het ministerie uitte. De minister besloot vervolgens niet in hoger beroep te gaan, maar de gevraagde inspectiegegevens aan Trouw te overhandigen (zie ook het hoofdstuk van Zoontjens).

Begin september werd ook via de media bekend dat Trouw het beroep gewonnen had. Deze publiciteit was voor Dronkers aanleiding om met het dagblad contact op te nemen. Hij wilde daarmee bereiken dat al in de eerste openbare publicatie van inspectiegegevens rekening gehouden zou worden met verschillen tussen scholen in leerlingenpopulatie en tussentijdse selectiviteit. Agerbeek worstelde ondertussen met de vraag hoe de verschillen tussen scholen het best uitgedrukt konden worden. Zij wilde op de eerste plaats dat het de

lezer duidelijk moest zijn dat rendementscijfers worden beïnvloed door de sociale achtergrond van leerlingen. Ten tweede wilde zij geen grote hoeveelheden cijfers over de lezers uitstorten. Zij zocht naar mogelijkheden om met eenvoudig rekenwerk naast de gegevens één gegeven te publiceren dat in één oogopslag een school typeert ten opzichte van de andere.

## *2.2 De gevolgde werkwijze*

In oktober 1997 maakte Trouw in een bijlage de schoolprestaties van alle scholen voor voortgezet onderwijs in Nederland openbaar (Agerbeek, Hageman, Kreulen & Lakmaker, 1997). Geordend naar provincie en schooltype publiceerde het dagblad voor elke school het percentage 'onvertraagd geslaagden' (het percentage leerlingen dat het eindexamen haalt zonder in het voorlaatste leerjaar te zijn blijven zitten), het percentage uitvallers en zittenblijvers op de school en het gemiddelde van de examencijfers voor Nederlands, Engels en Wiskunde A. Deze gegevens werden voor vbo, mavo, havo en vwo afzonderlijk weergegeven.

Naast deze ongecorrigeerde cijfers werd per school een rapportcijfer afgedrukt, dat was uitgerekend door Dronkers. In de bijbehorende toelichting staat: “het rapportcijfer voor de school brengt alle voorafgaande gegevens samen. Het zegt niet alleen iets over de prestaties van de leerlingen, maar ook over de bijdrage van de school daarin. Vandaar dat een school met een hoog percentage geslaagden een lager rapportcijfer kan hebben dan een school met een lager percentage. De oorzaak ligt dan in relatief lage eindexamencijfers, veel uitvallers en zittenblijvers of weinig allochtonen. (...) Een zes staat voor een gemiddelde prestatie. (...) Het cijfer houdt rekening met de mogelijkheden van de leerlingen als ze de school binnen komen. Ook het rendement van een school, dat wil zeggen hoeveel leerlingen aan een diploma zijn geholpen, hoe snel dat gaat en hoeveel leerlingen onderweg zijn uitgevallen, is meegewogen. Het gaat om een optelsom van het slagingspercentage, de eindexamencijfers voor Nederlands, Engels en wiskunde en de uitval- en zittenblijfpercentages tussen de brugklas en de examenklas. De uitkomst is van deze som is gecorrigeerd voor het percentage allochtonen op een school. De gegevens wegen niet even zwaar. Van elk berekende Dronkers de gemiddelde invloed op de examenresultaten en hield daar rekening mee. (...) Hoewel het percentage allochtonen volgens Dronkers een belangrijke indicator is voor de sociaal-culturele verschillen tussen leerlingen, is het zeker niet de enige of de beste. Zo zou het beter zijn als bekend was met welk Cito-toetsresultaat leerlingen de school binnenkomen. Maar dat wordt helaas nergens geregistreerd”.

Het rapportcijfer dat Dronkers voor elke school maakte is een cijfer voor het totale rendement van een schooltype binnen een school. Dit rendement is een optelsom van drie examencijfers, het slagingspercentage, het uitvalpercentage (twee typen) en het percentage zittenblijvers; de laatste twee voor elke leerjaar afzonderlijk, met uitzondering van de brugklas (geen goede gegevens beschikbaar) en de eindexamenklas (overlap met slagingspercentage). Zo gebruikte Dronkers voor vwo-, havo-, mavo- en vbo-afdelingen respectievelijk vijftien, dertien, tien en negen indicatoren. Omdat hij meer afzonderlijke indicatoren gebruikte dan in de Trouw-lijsten waren afgedrukt, kon niet rechtstreeks het cijfer uit deze lijsten worden afgeleid. De onderlinge gewichten van deze indicatoren bepaalde hij met een zogeheten factoranalyse per schooltype. Daarbij geldt dat het rapportcijfer hoger is naarmate de eindexamencijfers hoger zijn, het slagingspercentage hoger is en de percentages zittenblijvers en uitvallers lager zijn. Deze optelsom van rendementsindicatoren gaf het ongecorrigeerde rendement van een school weer.

Vervolgens bepaalde Dronkers in een zogeheten regressieanalyse de invloed van het percentage allochtone leerlingen (of beter: het percentage Cumi-leerlingen, zie paragraaf 3.1)

op het rendement van een school. Uit deze analyses bleek dat naarmate het percentage allochtone leerlingen op een school hoger is de rendementsscore lager is. Om dat verschil (dat te maken heeft met de kenmerken van de leerlingenpopulatie en niet met die van de school) te compenseren krijgen scholen naarmate ze meer allochtone leerlingen hebben een steeds grotere bonus. Uit deze berekening van het gecorrigeerde rendement komt per school een gestandaardiseerde score. Bij die score telde Dronkers zes punten op, zodat de gemiddelde school een zes krijgt. Scholen met een vijf of lager behoren tot het kwart laagst scorende scholen. Scholen met een zeven of meer maken deel uit van het kwart hoogst scorende scholen. Het schoolcijfer zegt volgens Dronkers (1997b: 12) niet dat scholen met een vijf onvoldoende functioneren: "De inspectie of overheid heeft nooit vastgesteld bij hoeveel zakkers, zittenblijvers of uitstromers een school onvoldoende functioneert". Als laatste stap heeft Dronkers de rapportcijfers afgerond op hele cijfers.<sup>2</sup> Wanneer er binnen een bepaald schooltype minder dan vijftig eindexamenkandidaten zijn, kreeg een school geen rapportcijfer.

De editie van Trouw met de schoolprestatielijst was binnen enkele uren volledig uitverkocht. Daardoor konden velen niet direct kennisnemen van de toelichting. Ook publiceerden veel regionale dagbladen deze rapportcijfers zonder nadere toelichting of nuancering. Het gevolg daarvan was dat bij velen de indruk ontstond dat de berichtgeving oppervlakkig en summier was.

### 3. Reacties

#### 3.1 Maatschappelijke discussie

Op het Trouw-onderzoek kwamen positieve en negatieve reacties. In de eerste weken na de publicatie stonden in vele dag- en weekbladen commentaren en ingezonden brieven. Ook op radio en televisie was er volop aandacht voor het onderzoek. De Trouw-redactie kreeg lof, omdat de resultaten van verschillende scholen nu eindelijk openbaar werden. Ook constateerde menigeen met vreugde dat de redactie niet alleen de slagingspercentages maar ook andere rendementsuitkomsten van scholen in de beoordeling betrok. Anderen brachten naar voren dat scholen nu beter konden worden gecontroleerd en juichten het toe dat het afgelopen was met de beschermde status van docenten. Een groot deel van de reacties was echter negatief over de publicatie. Op die negatieve reacties wordt hier ingegaan.

#### De inzichtelijkheid van het rapportcijfer

In veel reacties kwam naar voren dat het rapportcijfer *niet inzichtelijk* was. Van der Hilst (1997: 11) constateert dat het bijzonder vervelend is dat Trouw de berekening van het rapportcijfer niet vermeldt. "Als Trouw (of Dronkers) het te doen is om scholen een handvat te reiken in het evalueren van één van hun resultaten, is het toch te gek voor woorden dat scholen en hun consumenten niet op eenvoudige wijze deze berekening voorgeschoteld krijgen". Op de dag van publicatie was, naast de hierboven vermelde toelichting in Trouw, alleen een bewerkte SPSS-uitdraai beschikbaar bij Dronkers. Daarom publiceerde Dronkers (1997b) later op de opiniepagina van Trouw alsnog een meer uitgebreide verantwoording met een omrekenformule.<sup>3</sup>

#### De kwaliteit van de inspectiegegevens

Ook bleek dat sommige gegevens over scholen *foutief of twijfelachtig* van kwaliteit waren. Door gebruik van de term 'percentage allochtonen' in plaats van 'percentage Cumi-leerlingen' in de Trouw-publicatie zijn misverstanden gerezen (Louwerse, 1997). Een eerste

probleem is dat verscheidene leerlingen als autochtoon worden beschouwd, terwijl ze een andere etnische afkomst hebben. Dit geldt met name voor Surinaamse, Antilliaanse en Arubaanse leerlingen. Dit begrijpelijke bezwaar vloeit voort uit de definitie van Cumi-leerling. Leerlingen met de Nederlandse nationaliteit vallen buiten deze definitie.

Een tweede probleem met dit percentage was dat het niet per schooltype en ook niet per locatie, maar voor de gehele school (BRIN-nummer) gold. Die percentages waren een jaar later per schooltype beschikbaar. Dronkers (1998c) vergeleek vervolgens de uitkomsten met en zonder een uitsplitsing van het percentage Cumi-leerlingen naar schooltype. De uitkomsten verschillen van elkaar, maar de correctie verandert niet al te zeer (de samenhang in de via de twee methoden gecorrigeerde rapportcijfers was 0,62 voor vwo en 0,72 voor havo).

Ten derde waren categorale scholen met de registratie van het percentage zittenblijvers en uitvallers in het nadeel ten opzichte van brede scholengemeenschappen. Als een leerling na zittenblijven naar een ander schooltype binnen dezelfde school overstapt, dan registreert de inspectie dat doorgaans niet als uitval. Een overstap naar een andere school wordt echter wel als uitval opgevat. Soms was dat om formele redenen zelfs het geval binnen scholengemeenschappen. Het Sittards College scoorde voor de havo-, de atheneum- en de gymnasiumafdeling respectievelijk een zes, een vier en een zeven. In het Dagblad De Limburger (Redactie, 1997: 7) stelt de directie dat de school verkeerd is beoordeeld. "Scholieren van het College die intern van het gymnasium naar het atheneum gaan, of van het atheneum naar de havo, worden in Sittard aangemerkt als schoolverlaters. In het onderzoek van Trouw telt een schoolverlater als een minpunt, terwijl er in het geval van het College slechts sprake is van een interne verplaatsing. (...) Ik heb er geen moeite mee dat de onderwijsresultaten bekend worden gemaakt en dat er vergelijkingen worden getrokken. We hebben wel problemen met de selectieve en niet deskundige bewerking van die cijfers". Inderdaad zat in de inspectiegegevens een bias in het voordeel van brede scholengemeenschappen (zie ook Dronkers, 1998a).

#### Sociaal milieu en instroomniveau

Ook zou er onvoldoende rekening worden gehouden met het sociaal milieu of het instroomniveau van leerlingen. Het percentage Cumi-leerlingen dekt die verschillen in leerlingenpopulatie niet voldoende. Zoals in de toelichting al werd verteld beschikte de Inspectie en dus ook Trouw niet over die gegevens. Maar ook als die gegevens er wel waren geweest, zouden de rapportcijfers volgens Dronkers niet met grote sprongen zijn veranderd. Veenstra, Dijkstra, Peschar en Snijders (1998) laten namelijk zien dat het percentage allochtone ouders matig negatief samenhangt met de gemiddelde sociaal-economische status en het gemiddelde aanvangsniveau van leerlingen op een school. Voor scholen met veel laaggeschoolde autochtone leerlingen maar met weinig allochtone leerlingen (bijvoorbeeld op het platteland of in achterstandswijken buiten de Randstad) is echter wel onvoldoende rekening gehouden met het sociaal milieu of het instroomniveau van de leerlingen.

#### Bijzondere omstandigheden van scholen

Verscheidene schoolleiders gaven aan dat de cijfers geen rekening houden met de *bijzondere omstandigheden* waarin sommige scholen verkeren. Zo heeft het Trouw-onderzoek geen rekening gehouden met de aanwezigheid van een ivbo-afdeling of een internationale schakelklas (ISK). De plaatsvervangend rector van een slecht scorende school heeft in het NRC Handelsblad (Redactie Onderwijs, 1997: 3) een verklaring voor het lage slagingspercentage van zijn vbo-afdeling met veel leerlingen die de individuele route volgen: " We hebben ook de enige ISK in Groningen, met instromende buitenlandse kinderen. Zij

doen er gemiddeld ook veel langer over, maar daar is geen rekening mee te houden. De kille cijfers betekenen dus niet dat we een slechte school zijn. Dat kan niet, dat zijn we niet”.

Bij de beantwoording van de inspectieformulieren hadden scholen zelf leerlingen met bijzondere kenmerken (zoals ISK en asielzoekers) buiten beschouwingen moeten laten. In een aantal gevallen hadden scholen dat echter verzuimd.

#### Getallen over verscheidene jaren

Om te kunnen spreken van kwaliteit van scholen, zo haalde ook menigeen aan, moet er op zijn minst sprake zijn van cijfers over verscheidene schooljaren. Het moet namelijk niet zo zijn dat toevallige omstandigheden die met het schoolbeleid en het functioneren van de leerlingen niets te maken hebben van invloed zijn op de schoolprestatie. Dit bezwaar lijkt redelijk, maar Trouw had slechts om gegevens van één schooljaar gevraagd. In een vervolganalyse heeft Dronkers (1998b) laten zien dat de stabiliteit van de belangrijkste indicatoren van het intern rendement van scholen (samenstelling leerlingenpopulatie, eindexamencijfers voor het centraal schriftelijk examen; slaagpercentage) matig tot hoog is.

#### Andere kwaliteitsindicatoren

Het Trouw-onderzoek meet volgens velen überhaupt niet de *kwaliteit van het onderwijs*. Ook andere zaken kunnen van belang zijn, zoals de creatieve en de sociaal-emotionele ontwikkeling van leerlingen of de mate waarin ze zelfstandig kunnen leren. Agerbeek e.a. (1997) kwam in de toelichting bij de schoolprestatelijsten al bij voorbaat tegemoet aan dergelijke bezwaren: "Kwaliteit is een subjectief begrip. Zo vinden sommige ouders de identiteit van de school belangrijker dan het slagingspercentage, hechten anderen vooral aan de afstand van huis tot school, laten weer anderen zich leiden door de goede naam van een school en zijn er ook ouders die rekening houden met het pedagogisch concept. En gelijk hebben ze. Een cijfer is niet meer dan een van de criteria waarop een school kan worden beoordeeld. De schoolprestatelijst is dan ook een steuntje in de rug bij de schoolkeuze, niets meer en niets minder”.

Dronkers en Warnaar (1999) hebben de gegevens uit het Trouw-onderzoek gecombineerd met data uit het Nationaal Scholierenonderzoek. Ze lieten zien dat de verschillende dimensies van kwaliteit van scholen (intern rendement; extern rendement; culturele vorming; welzijn van scholieren) niet haaks op elkaar staan, maar eerder positief samenhangen.<sup>4</sup> Een school die het goed doet op de ene kwaliteitsindicator, doet het doorgaans ook goed op de andere. Er lijkt over het algemeen geen sprake van een compensatie-effect. Ook wijst Dronkers (1998a) er op dat indicatoren voor die andere kwaliteiten van scholen bij de inspectie ontbreken (zie voor een overzicht van mogelijke kenmerken: Landelijk Actie Komitee Scholieren, 1996). Blijkbaar hebben de politiek en de onderwijsorganisaties het nooit de moeite waard gevonden over die andere belangrijke dimensies gegevens te laten verzamelen.<sup>5</sup>

### 3.2 Wetenschappelijke discussie

In de meer wetenschappelijke discussie naar aanleiding van de Trouw-publicaties kwamen, naast de hierboven vermelde zaken, ook nog andere onderwerpen aan de orde.

#### Meerniveau analyse

Veenstra e.a. (1998) laten zien dat de keuze van de analysemethode gevolgen heeft voor de scores van scholen. Scholen die met een op schoolniveau geaggregeerde regressieanalyse (de methode van het Trouw-onderzoek) bijvoorbeeld een zes halen, kunnen met een meerniveau analyse zo maar een punt lager of hoger scoren.<sup>6</sup> In het hoofdstuk van Bosker, Béguin en Rekers-Mombarg worden vergelijkbare verschillen tussen analysemethoden aangetoond.

Veenstra e.a. concluderen dat een eerlijke vergelijking van scholen onmogelijk is zo lang er van alle scholen geen gegevens voor een meerniveau analyse beschikbaar zijn (zie ook: Bosker, Lam, Luyten, Steen & De Vos, 1998; Bosker & Scheerens, 1999).

In een reactie op Veenstra e.a. (1998) vindt Dronkers (1998a) echter dat met de keuze voor die analysemethode te hoog wordt ingezet wat de kwaliteit van data betreft. Aangezien er voorlopig geen meerniveau gegevens zijn van alle scholen voor voortgezet onderwijs, is meerniveau analyse volgens Dronkers (1998a: 149) niet toepasbaar: "Als wetenschappers dan toch staan op meerniveau analyses, sluiten zij zichzelf uit van het verdere maatschappelijke debat over de meting van 'toegevoegde waarde' van scholen. De wetenschappelijke opvattingen over de favoriete methode hebben dan de bovenhand gekregen, waardoor de maatschappelijke dienstverlening door de wetenschap gehinderd zou worden. In dat geval is het mogelijk betere (een superieur geachte meerniveau methode) de vijand van het goede (een min of meer verantwoorde meting van kwaliteit van scholen)". Elders in dat artikel stelt Dronkers (1998a: 143): "Wetenschappers hebben naast hun wetenschappelijke verantwoordelijkheid ook een maatschappelijke verantwoordelijkheid. Dat laatste betekent ook dat wetenschappers bereid moeten zijn geaccepteerde wetenschappelijke inzichten toe te passen op maatschappelijke vraagstukken, juist als dat maatschappelijk nog niet geaccepteerd is. (...) Het berekenen van het rapportcijfer is daarom geen sociaal-wetenschappelijk onderzoek, maar maatschappelijke dienstverlening met behulp van wetenschappelijke inzichten". Als aanzet tot een rangordening van scholen mag deze maatschappelijke dienstverlening geslaagd worden genoemd. Het blijft echter betwistbaar of deze manier van rangordenen wetenschappelijk gezien goed genoeg is.

Een goede school op grond van de Trouw-criteria behoeft niet voor alle leerlingen een goede school te zijn. Bosker en Scheerens (1999) wijzen op het belang van differentiële opbrengsten van scholen, bijvoorbeeld dat een bepaalde school vooral geschikt is voor leerlingen met leerachterstanden en minder geschikt is voor de betere leerlingen. Dronkers (1999a) ontkent niet het bestaan van deze differentiële effecten, die alleen met een meerniveau aanpak zichtbaar kunnen worden. Hij stelt dat dit voorbehoud bij elke Trouw-publicatie is gemaakt. Hij vindt echter dat het belang van deze differentiële effecten niet overdreven mag worden: slechts bij een beperkt aantal scholen treden deze effecten op, zoals blijkt uit de analyses van Bosker en Scheerens (zie ook het hoofdstuk van Bosker, Béguin en Rekers-Mombarg).

### Longitudinale gegevens

Van der Werf en Creemers (1999) pleiten voor longitudinaal gemeten rendementen in plaats van de cross-sectioneel gemeten rendementen. Aan het gebruik van longitudinale data voor de publieke openbaarmaking van schoolrendementen kleven volgens Dronkers (1999a) echter bezwaren. Longitudinale gegevens over alle scholen zijn er op dit moment nog niet. Bovendien gaat het bij longitudinale gegevens vooral om het functioneren van scholen van minimaal vier tot zes jaar geleden. Met longitudinale gegevens kan pas een uitspraak over het functioneren van een school worden gedaan, wanneer de laatste zittenblijvers van een bepaald cohort de school hebben verlaten. Wanneer een school dan een laag rapportcijfer haalt, kan de directie zeggen dat dat komt door problemen van een aantal jaren terug, maar dat die problemen inmiddels overwonnen zijn. "Kortom, het openbaar publiceren van prestatiegegevens uitsluitend op basis van longitudinale gegevens komt voor de externe controle, voor het kiezen van ouders en voor schoolverbetering als mosterd na de maaltijd (Dronkers, 1999a: 78)".

## Factoranalyse

Bosker verwerpt de factoranalyse, omdat de computer dan bepaalt *in welke mate* factoren, zoals het percentage geslaagden, zittenblijvers, uitvallers en eindexamencijfers bijdragen aan de beoordeling van een school (Bosker in Jager, 1998: 14). "Dat is alsof je bij een WK schaatsen alle tijden op vier afstanden in de computer stopt. Vervolgens zeg je tegen de deelnemers: 'Wacht even, de computer bepaalt nu hoe zwaar elke afstand voor de uitslag meetelt, en dan weten we zo wie wereldkampioen is'. Als daar dan de 5000 meter als belangrijkste uitrolt is dat vreemd, want bij het vorige kampioenschap was dat nog de 1500 meter. Aan zo'n koffiemolentechniek kun je een dergelijke beslissing niet toevertrouwen".

Uit een analyse van Dronkers (1998b) blijkt inderdaad dat de gewichten van jaar tot jaar verschillen. Hij pleit er daarom voor om de gewichten van de verschillende indicatoren op normatieve gronden te kiezen in plaats van ze op grond van een aantal schooljaren vast te stellen (Dronkers, 1998b; 1999a). Zo zwart als Bosker de stabiliteit van de factorscores schetst is het echter niet. De eindexamencijfers en de slagingspercentages van twee opeenvolgende jaren hangen matig positief samen. Het zijn de gewichten voor de uitstroom- en zittenblijpercentages die instabiel zijn.<sup>7</sup>

## Tussenschoolse variantie

De suggestie moet niet zijn dat verschillen in prestaties van leerlingen vooral op het conto van een school moeten worden geschreven. Uit onderzoek (Veenstra, 1999) blijkt dat de verklaring voor verschillen in prestaties van leerlingen niet zo zeer moet worden gezocht in verschillen tussen scholen, maar vooral in verschillen binnen scholen. Wat die verschillen binnen scholen betreft, moet worden gedacht aan verschillen in bijvoorbeeld de gelegenheid tot leren tussen klassen, de mate van gestructureerd lesgeven door docenten, of de intelligentie en de gezinsachtergrond van leerlingen.

De rapportcijfers uit de Trouw-publicatie suggereren een enorme kwaliteitsvariatie, maar dat is inherent aan de methode (Bosker & Scheerens, 1999). Wanneer scores worden gestandaardiseerd met een zes als gemiddelde en een standaarddeviatie van één dan heeft automatisch 40 procent van de scholen een score tussen 5,5 en 6,5. Deze scholen krijgen dus een zes als cijfer. Van de overige scholen scoort ongeveer 30 procent lager dan een 5,5 en een zelfde percentage een 6,5 of hoger. Hoe gering de verschillen binnen een schooltype in werkelijkheid ook zijn, er zal op deze manier altijd blijken dat er verschillen tussen scholen zijn.

Uit analyses van Veenstra (1999) blijkt echter dat uitgaande van honderd eindexamenleerlingen binnen een schooltype slechts één op de vier scholen significant afwijkt van het gemiddelde. Van deze 25 procent wijkt de helft positief en de andere helft negatief af.

## 4. Het vervolg

Eind 1997 wijzigde de staatssecretaris haar beleid onder invloed van de Trouw-publicatie. Zij concludeerde toen dat er kennelijk behoefte bestond aan gesystematiseerde overzichten van resultaatgegevens van scholen. Daarom besloot zij dat de inspectie jaarlijks rapportages van de kwaliteit van afzonderlijke scholen moet uitbrengen: de 'kwaliteitskaart' (brief van Netelenbos aan de vaste commissie van OCW, 19 december 1997). "Bij de ontwikkeling van deze kwaliteitskaart zal geleidelijkheid noodzakelijk zijn: niet alle gegevens zijn direct beschikbaar en er moet ervaring opgedaan worden met manieren om kwaliteit in beeld te brengen. Om de gegevens van individuele scholen in een breder verband te kunnen plaatsen,

zal de Inspectie een typologie van scholen ontwikkelen, waarmee categorieën van scholen kunnen worden onderscheiden die werken met vergelijkbare instroom en omstandigheden. (...) Ook in het basisonderwijs start de inspectie met een nieuwe werkwijze, die zij aanduidt als 'de integrale schoolevaluatie'. De bedoeling is ook hier te komen tot een completer beeld van de kwaliteit van de afzonderlijke school, rekening houdend met kenmerken van (de samenstelling van de leerlingenpopulatie op) de school. De ervaringen die opgedaan worden met deze werkwijze kunnen tot nader inzichten leiden met betrekking tot de weergave van resultaten in de schoolgids. De voorgestelde algemene maatregel van bestuur opent daartoe de mogelijkheid".

#### *4.1 Kwaliteitskaart*

De Onderwijsinspectie heeft in het najaar van 1998 voor het eerst een kwaliteitskaart voor scholen voor voortgezet onderwijs uitgebracht. Per school staan op deze kaart voor alle schooltypen van elke locatie kerngegevens vermeld. Het gemiddelde examencijfer is een van die gegevens. Dit cijfer wordt op zes manieren weergegeven, namelijk voor alle vakken en afzonderlijk voor Nederlands, talen (Engels, Frans en Duits), aardrijkskunde en geschiedenis, economische vakken en bètavakken (biologie, natuurkunde, scheikunde, wiskunde). Voor de eerste drie leerjaren wordt aangegeven hoeveel procent van de leerlingen niet blijft zitten. Voor de hogere leerjaren, dat wil zeggen vanaf de derde klas, wordt berekend welk percentage leerlingen zonder vertraging een diploma haalt. Ook vermeldt de kaart hoe groot de klassen in de eerste twee leerjaren zijn.

De kwaliteitskaart geeft ook indicaties van de toegevoegde waarde van de scholen. Voor de berekening daarvan gebruikt de inspectie het percentage Cumi-leerlingen, de urbanisatiegraad van de gemeente waarin de school staat en het al dan niet openbaar zijn van de school. Deze toegevoegde waarde van het intern rendement wordt niet in één cijfer samengevat, maar voor elke indicator afzonderlijk vastgesteld.<sup>8</sup>

De Onderwijsraad (1998), het adviesorgaan voor de regering op het terrein van onderwijs, heeft gesteld dat de overheid zich bij het berekenen van kwaliteit moet baseren op controleerbare, objectieve gegevens. Om dat te bereiken zouden alle leerlingen een aanvangstoets moeten maken. De resultaten op die toets zouden dan kunnen worden vergeleken met de prestaties aan het eind van het voortgezet onderwijs. Zo lang een dergelijke begintoets er nog niet is, moet de overheid volgens de Onderwijsraad niet een toevlucht nemen tot vage, onduidelijke gegevens, bijvoorbeeld het percentage allochtone leerlingen, de urbanisatiegraad of de richting van de school. De leden van dit adviesorgaan vinden dus dat de kwaliteitskaart van het ministerie aan hogere standaarden moet voldoen dan de beoordeling door Trouw. Verder vindt de Onderwijsraad dat het rendement van scholen op de kwaliteitskaart centraal moet staan en dat alleen zaken moeten worden opgenomen die wettelijk zijn vastgelegd, de zogeheten deugdelijkheids-eisen (zie ook het hoofdstuk van Van Dyck en Leune).

In de kwaliteitskaart ontbreekt een samenvattend oordeel, zoals het rapportcijfer van het Trouw-onderzoek. Dronkers (1997a) protesteerde tegen deze omissie. "Ouders zullen met de vele kengetallen op de kaart al heel snel door de bomen het bos niet meer kunnen zien. Want scholen met een slechte score op het ene kengetal (bijv. een hoog percentage zittenblijvers) zullen snel wijzen op een andere, voor hen gunstiger uitkomst (bijv. een hoog eindexamencijfer voor een eindexamenvak). En als ook die ontbreken kunnen scholen nog altijd wijzen op de samenstelling van de leerlingenpopulatie". Verder vond Dronkers het gebruik van de richting van de school als correctiefactor bij de berekening van de toegevoegde waarde strijdig met de grondwettelijke gelijkheid van scholen in Nederland.



#### *4.2 Nieuwe versies van de kwaliteitskaart*

De kwaliteitskaart van de onderwijsinspectie vermeldt vanaf het schooljaar 1999-2000 niet alleen het uitstroomniveau (eindexamencijfers), maar ook het instroomniveau (advies voor voortgezet onderwijs). Door zowel het instroom- als uitstroomniveau weer te geven, is volgens de Inspectie te zien wat leerlingen op een school leren en of een school meer of minder dan gemiddeld bijbrengt. De speciale omstandigheden waaronder een school werkt, hoeven daardoor niet meer te worden meegenomen. Er wordt dus niet meer gecontroleerd voor het percentage Cumi-leerlingen, de urbanisatiegraad of de richting van de school. De Inspectie volgt hiermee ten dele de kritiek van de Onderwijsraad (1998) op. Het advies voor voortgezet onderwijs is namelijk een tamelijk heldere maat voor de samenstelling van de leerlingenpopulatie.

Het advies voor voortgezet onderwijs is echter niet ideaal om het instroomniveau van leerlingen vast te stellen. Allochtone leerlingen en leerlingen uit hogere sociale milieus krijgen gemiddeld genomen hogere adviezen dan hun score op de Cito-toets aangeeft ('t Hof & Dronkers, 1993; Koeslag & Dronkers, 1994). Ook zijn er verschillen in advies naar regio en denominatie (Dronkers, Van Erp, Robijns & Roeleveld, 1998). Wanneer de enige school voor voortgezet onderwijs van een bepaalde denominatie in een bepaalde regio alleen maar dakpansgewijze brugklassen kent, dan geven de basisscholen van diezelfde denominatie bijpassende adviezen. Zo kan het dus gebeuren dat voor leerlingen op een bepaalde school het havo/vwo-advies het hoogst haalbare advies is. Wanneer een deel van deze leerlingen vervolgens doorstroomt naar het vwo, dan scoren deze leerlingen zogenaamd boven het advies van hun basisschool en is de toegevoegde waarde van hun school zogenaamd groot. Om dergelijke gebreken in de gegevensverzameling tegen te gaan en het rendement van een school eerlijk te meten blijft een aanvangstoets het ideaal.

#### *4.3 Nieuwe versies van de Trouw-publicatie*

Van de Trouw-publicatie zijn inmiddels drie nieuwe versies verschenen. De berekening in de tweede editie van de Trouw-publicatie (Agerbeek, 1998) leek sterk op het eerdere rapportcijfer. In de tweede editie liepen de beoordelingen niet van één tot en met tien, maar werden de mogelijke scores uitgedrukt in plussen en minnen. De beide beoordelingen zijn om te rekenen: een zes als rapportcijfer is bijvoorbeeld een nul als beoordeling en een vier of lager als rapportcijfer komt overeen met een dubbele min. Dronkers (1998b) analyseerde de overeenkomsten en verschillen tussen de methoden en uitkomsten van beide publicaties. Voor de tweede editie van het Trouw-onderzoek heeft het dagblad haar analyses gebaseerd op betere gegevens van scholen. Scholen hebben de inspectieformulieren nauwkeuriger en op een meer vergelijkbare wijze ingevuld. Zo worden de uitkomsten van leerlingen die buiten het reguliere onderwijs vallen, zoals leerlingen in internationale schakelklassen, niet langer ten onrechte door scholen meegerekend. Voor de cijfers geldt bovendien dat ze beschikbaar zijn per afzonderlijke locaties en niet alleen per school. Het percentage Cumi-leerlingen was nu ook per onderwijstype per school beschikbaar, maar nog niet per locatie. In deze versie werd het percentage bètavakken voor vwo en havo en het percentage vakken op D-niveau voor mavo en vbo als indicator voor het intern rendement opgenomen. De analyses laten zien dat het al dan niet opnemen van deze indicator weinig uitmaakt voor het rapportcijfer van een school.

In de derde (1999) en vierde (2000) versie heeft Trouw de gegevens van de kwaliteitskaarten van de Inspectie gebruikt. Met deze gegevens heeft het dagblad wederom een beoordeling met plussen en minnen opgesteld. Deze beoordeling is gebaseerd op de gemiddelde inspectiebeoordeling van drie elementen, te weten het aantal leerlingen dat onvertraagd in de derde klas belandt, het aantal leerlingen dat in de bovenbouw zonder zitten te

blijven het diploma haalt en het gemiddelde eindexamencijfer. Daarbij is rekening gehouden met het advies van leerlingen voor voortgezet onderwijs (en voor het rendement in de bovenbouw ook met het percentage Cumi-leerlingen en het percentage leerlingen met een tegemoetkoming in de studiekosten).

Agerbeek (1999a) publiceerde naast deze derde en vierde versie van het Trouw-onderzoek al twee keer een eigen keuzegids, waarin zij veel gegevens over scholen samenvat.

## 5. Conclusie en discussie

Het Trouw-onderzoek kan worden beschouwd als een eerste aanzet tot het voor een breder publiek zichtbaar maken van de prestaties van scholen. Het heeft de weg gebaad voor een betere verzameling en bewerking van inspectiegegevens en een systematische vaststelling van het interne rendement van scholen door de verantwoordelijke overheid, waarbij rekening wordt gehouden met de leerlingenpopulatie, het instroomniveau en de tussentijdse selectie binnen scholen.

De Trouw-publicatie en de kwaliteitskaart van de Inspectie hebben tot nu toe alleen het intern rendement van scholen onderling vergeleken, zonder daarbij echter een grens tussen voldoende en onvoldoende aan te geven. Toch is deze grens voor beleidsmakers, toezichthouders, scholen en ouders van belang.

Ouders lijken rekening te houden met de Trouw-beoordelingen en de kwaliteitskaarten, maar laten zich er niet uitsluitend door leiden (Agerbeek, 1999b; Dronkers, 1999b). Ouders kiezen vaker scholen met een hoog cijfer dan zij in de voorafgaande jaren deden. Deze invloed is overigens niet groot. Van alle waargenomen verschillen in het groeipercentage van scholen kan 2 procent van de variantie worden verklaard door dat cijfer. Ouders hebben zich niet laten beïnvloeden door de gepubliceerde sociale compositie (percentage Cumi-leerlingen) van een school (zie ook het hoofdstuk van Dijkstra en Witziers). Verder suggereren de resultaten dat ouders zich bij de keuze van een school weinig aantrekken van gemeentegrenzen. Verscheidene ouders laten hun kinderen voortgezet onderwijs volgen in een andere plaats dan hun woongemeente.

In de eerste Trouw-publicatie kwamen echter ook gebreken voor. Zo zou het beter zijn geweest om de verantwoording gelijk af te drukken in plaats van veertien dagen later; om duidelijker uit te leggen dat het rapportcijfer niet rechtstreeks uit de gepubliceerde gegevens was af te leiden; om de rapportcijfers om te zetten in plussen en minnen; om meer te benadrukken dat de inspectiegegevens niet al te betrouwbaar zijn, doordat scholen deze vaak onnauwkeurig invulden; om het gemiddelde percentage zittenblijvers en uitvaller per school en locatie te gebruiken in plaats van deze percentages per leerjaar; om bij de berekening van het totale interne rendement een zelfde gewicht voor de verschillende rendementsindicatoren te gebruiken; en om in de toelichting duidelijk aan te geven dat verschillen in prestaties van leerlingen slechts in geringe mate op het conto van een school kunnen worden geschreven.

Een deel van deze minder geslaagde keuzen vloeide voort uit de randvoorwaarden waaronder de eerste Trouw-publicatie geproduceerd moest worden (bijvoorbeeld het rapportcijfer van één tot tien). Een ander deel van de minder geslaagde keuzen vloeide voort uit een te grote voorzichtigheid (geen middeling van de zittenblijf-percentages). Een laatste deel hing samen met een te groot vertrouwen in de kwaliteit van de inspectiegegevens en de ernst waarmee scholen zich van hun wettelijke taak kweten. Blijft staan dat de vaststelling van de effecten van scholen het zuiverst gebeurt met meerniveau analyse. Voor de realisatie van deze wens is het wachten op de invoering van het leerlingnummer.

## Literatuur

- Agerbeek, M. (1998). Schoolprestaties 1998. *Trouw*, 13 juni, ZENZ.
- Agerbeek, M. (1999a). *Keuzegids middelbare scholen 1999*. Zeist: Kerckebosch.
- Agerbeek, M. (1999b). Wie geschoren wordt, moet stilzitten. *Trouw*, 11 mei, 15.
- Agerbeek, M., Dronkers, J., Ormel, J. & Vermeulen, M. (2000). De publicatie van schoolinspectiecijfers door Trouw: een goed voorbeeld van de complexe relaties tussen journalistiek en wetenschap. In P. Ester, D. van Houten & B. Steijn (eds.), *De Waan van de Dag. Sociale wetenschap en de publieke zaak* (pp. 125-143) Amsterdam: SISWO.
- Agerbeek, M., Hageman, E., Kreulen, E. & Lakmaker, H. (1997). Trouw-onderzoek schoolprestaties. *Trouw*, 25 oktober, ZENZ.
- Bosker, R.J., Lam, J.F., Luyten, H., Steen, R. & Vos, H. de (1998). *Het vergelijken van scholen*. Enschede: Universiteit Twente.
- Bosker, R.J. & Scheerens, J. (1999). Openbare prestatiegegevens van scholen; nuttigheid en validiteit. *Pedagogische Studiën* 76, 61-73.
- Dronkers, J. (1997a). Netelenbos komt met volstrekt onvoldoende kwaliteitskaart van scholen. *Trouw*, 31 december, 11.
- Dronkers, J. (1997b). Schoolcijfer zegt niet dat scholen met 5 onvoldoende functioneren. *Trouw*, 8 november, 12.
- Dronkers, J. (1998a). Het betere is de vijand van het goede. Een reactie op de commentaren over het Trouw rapportcijfer. *Pedagogische Studiën* 75, 142-150.
- Dronkers, J. (1998b). Het Trouw-rapportcijfer van scholen in het voortgezet onderwijs; een analyse van de inspectiegegevens over de schooljaren 1995/96 en 1996/97. *Tijdschrift voor Onderwijsresearch*, 23, 159-176.
- Dronkers, J. (1998c). *Verschildt de toegevoegde waarde van het intern rendement van scholen in het voortgezet onderwijs? Enkele kanttekeningen bij het Trouw-onderzoek Schoolprestaties 1998*. Amsterdam: SCO-Kohnstamm Instituut.
- Dronkers, J. (1999a). Het betere blijft nog steeds de vijand van het goede. Een antwoord op Van der Werf en Creemers, en Bosker en Scheerens. *Pedagogische Studiën* 76, 74-81.
- Dronkers, J. (1999b). Veranderden leerlingaantallen in het voortgezet onderwijs in het schooljaar 1998-1999 door de publicatie van inspectiegegevens en de berekening van het schoolcijfer door Trouw in oktober 1997? Een nadere analyse. *Tijdschrift voor Onderwijsresearch*, 24, 63-66.
- Dronkers, J., Erp, M. van, Robijn, M. & Roeleveld, J. (1998). Krijgen leerlingen in de grote steden en met name in Amsterdam te hoge adviezen? *Tijdschrift voor Onderwijsresearch*, 23, 17-30.
- Dronkers, J. & Warnaar, M.F. (1999). 'Ze leren er niks maar er is een waanzinnig goede sfeer op deze school'. Over de samenhang tussen kwaliteiten van scholen. *VBSchrift*, 25, 7-16.
- Hilst, B. van der (1997). Montessori-scholen doen het beter dan hun omgeving. *Trouw*, 6 november, 11.
- Hof, L. van 't & Dronkers, J. (1993). Onderwijsachterstanden van allochtonen: klasse, gezin of etnische cultuur? *Migrantenstudies*, 9, 2-25.
- Jager, P. K. (1998). Kwaliteitskaart verdoezelt waarheid. *Het Onderwijsblad*, 27 juni 1998, 14.
- Koeslag, M. & Dronkers, J. (1994). Overadvisering en de schoolloopbanen van migrantenleerlingen en autochtone leerlingen in het voortgezet onderwijs. *Tijdschrift voor Onderwijsresearch*, 19, 240-258.

Landelijk Actie Komitee Scholieren (1996). *Eindrapportage Gouden School Enquête*. Amsterdam: LAKS.

Louwerse, P. (1997). De venijnige wraak op professor Dronkers. *Rotterdams Dagblad*, 10 december, 5.

Onderwijsraad (1998). *Kwaliteitskaart voortgezet onderwijs*. Den Haag: Onderwijsraad.

Putten, C.M. van & Kamp, L.J.T. van der (1998). Dronkers' rapportcijfer voor de middelbare scholen: nauwkeurigheid, geldigheid en betekenis. *Pedagogische Studiën* 75, 135-141.

Redactie (1997). Sittard College eist rectificatie Trouw. *Dagblad De Limburger*, 29 oktober, 7.

Redactie onderwijs (1997). Kille cijfers kloppen maar komen hard aan. *NRC Handelsblad*, 27 oktober, 3.

Veenstra, D.R., Dijkstra, A.B., Peschar, J.L. & Snijders, T.A.B. (1998). Scholen op rapport. Een reactie op het Trouw-onderzoek naar schoolprestaties. *Pedagogische Studien*, 75, 121-134.

Veenstra, R. (1999). *Leerlingen - klassen - scholen. Prestaties en vorderingen van leerlingen in het voortgezet onderwijs*. Amsterdam: Thela Thesis.

Werf, M.P.C. van der & Creemers, B.P.M. (1999). Naar een vollediger systeem van kwaliteitsbeoordeling. *Pedagogische Studiën* 76, 51-60.

---

<sup>1</sup> In deze paragraaf wordt gebruik gemaakt van Agerbeek, Dronkers, Ormel & Vermeulen (1999).

<sup>2</sup> Dronkers heeft de cijfers naar eigen zeggen afgerond om de indruk van exactheid te voorkomen. Die indruk moet met deze cijfers inderdaad niet worden gewekt. De keuze om dan maar af te ronden is achteraf echter niet gelukkig, omdat kleine verschillen tussen scholen dan groter kunnen lijken dan ze werkelijk zijn.

<sup>3</sup> In de afgedrukte omrekenformule voor de mavo bleek een zetfout te zitten. De gecorrigeerde versie van de mavo-omrekenformule is gepubliceerd in Dronkers (1998b).

<sup>4</sup> De analyses van Dronkers en Warnaar (1999) zijn verricht op geaggregeerd niveau.

<sup>5</sup> Van der Werf en Creemers (1999) wijzen er hierbij op dat het betrekken van procesindicatoren alleen zin heeft als zij een significante bijdrage leveren aan het verklaren van verschillen in schoolprestaties. Verder onderzoek zal volgens hen uit moeten maken welke proceskenmerken relevant zijn om in een beoordelingssysteem opgenomen te worden.

<sup>6</sup> Dronkers (1999a) betwijfelt of deze verschillen in resultaten geheel voor rekening komen van de gebruikte methode. Veenstra e.a. (1998) hebben namelijk een andere afhankelijke maat, de leerjarenladder, gebruikt en bovendien hebben ze leerlingen ook na het verlaten van een school gevolgd (zo nemen ze dus het extern rendement van een school mee). Hun criterium was de plaats op de leerjarenladder acht jaar na de intrede in het voortgezet onderwijs.

<sup>7</sup> Van Putten en Van der Kamp (1998) wijzen erop dat de door Dronkers geformuleerde interne rendementsmaat niet de enige of in enkele gevallen zelfs niet de belangrijkste factor is die uit de factoranalyse rolt. Hierbij wijzen ze er ook op dat de eerste factor steeds maar een eenvierde tot een eenvijfde van de variantie in de variabelen bindt. In een vervolganalyse heeft Dronkers (1999b) voor havo- en vwo-scholen de andere dimensies aan een nader onderzoek onderworpen. Waar Dronkers voor het Trouw-onderzoek een ongeroteerde factoranalyse uitvoerde, voert hij in zijn vervolganalyse wel een rotatie uit. In de geroteerde oplossing vormen slagingspercentages, eindexamencijfers en zittenblijf- en uitstroompercentages in het voorlaatste leerjaar de ene dimensie en zittenblijf- en uitstroompercentages in de lagere leerjaren de andere dimensie. In de tweede Trouw-publicatie heeft Dronkers (1999a) gewerkt met het gemiddelde percentage zittenblijvers en uitstromers. Dan blijkt dat de verschillende indicatoren duidelijk een factor vormen. De factorladingen zijn absoluut gezien 0,60 of hoger (met als uitzondering het gemiddelde eindexamencijfer op het vbo dat een lading van 0,41 heeft). Verder heeft Dronkers voor vwo en havo het percentage bètavakken en voor mavo en vbo het percentage examens op D-niveau meegenomen in zijn berekening van het intern rendement. De factorladingen hiervan variëren tussen de 0,32 en 0,37 (met als uitzondering het percentage examens op D-niveau voor de mavo dat een

---

lading van 0,70 heeft). Deze extra indicatoren hangen minder sterk samen met de maat voor het intern rendement.

<sup>8</sup> De gewichten van de factoren, gebruikt bij de berekening van deze verschillende toegevoegde waarden, werden met behulp van regressieanalyse door Bosker berekend.