

Discussie

Scholen op rapport.

Een reactie op het Trouw-onderzoek naar schoolprestaties

D.R. Veenstra, A.B. Dijkstra, J.L. Peschar en T.A.B. Snijders,
Rijksuniversiteit Groningen

1Inleiding

In het najaar van 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' (een schatting van het percentage leerlingen dat het eindexamen haalt zonder in het voorlaatste leerjaar te zijn blijven zitten), het percentage uitvallers en zittenblijvers op de school in de jaren daarvoor en het gemiddelde van de examencijfers voor Nederlands, Engels en Wiskunde A. Deze gegevens werden voor VBO, MAVO, HAVO en VWO weergegeven. Met deze gegevens berekende de onderwijssocioloog prof.dr J. Dronkers ten slotte een rapportcijfer. Behalve de voorgaande prestatie-maten werd in die berekening ook het percentage allochtonen betrokken. Door dit percentage mee te wegen probeerde *Trouw* te voorkomen dat scholen met een zwarte leerlingenpopulatie als vanzelf tot de slechtste worden gerekend.

Het ministerie van Onderwijs, Cultuur en Wetenschappen weigerde *Trouw* aanvankelijk de gegevens over schoolprestaties te verstrekken. De weigering was gebaseerd op twee gronden. Het doorgeven van het materiaal aan derden zou de vertrouwensrelatie van de onderwijsinspectie met de scholen aantasten. Ten tweede zouden de scholen schade ondervinden van publicatie van de gegevens omdat slagingspercentages geen rekening houden met de achtergrond van leerlingen. De redactie van het dagblad ging vervolgens naar de rechter en deed een beroep op de Wet Openbaarheid van Bestuur. De rechter stelde de krant 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.

In Engeland en Frankrijk zijn schoolprestaties openbaar en bestaan er al langer 'hitlijsten' van scholen. Per onderwijsregio publiceert de Engelse overheid jaarlijks de resultaten van scholen voor voortgezet onderwijs. In die publicatie staan twee maten centraal, het percentage leerlingen dat een minimum aantal certificaten voor voortgezet onderwijs (*general certificate of secondary education*) heeft behaald en de score van leerlingen op de bijbehorende toetsen. Verscheidene dagbladen zetten deze cijfers om in *league tables*, waarin de prestaties van scholen op een rij worden gezet. Het afgelopen jaar verscheen ook een overzicht van scholen waar de prestaties de laatste jaren het meest zijn verbeterd. Behalve de resultaten van het voortgezet onderwijs geeft de Engels overheid sinds vorig jaar ook die van basisscholen weer.

De kloof tussen goede en slechte scholen kan door dergelijke hitlijsten groter worden (vergelijk Dijkstra, 1996). Zo zou kunnen worden verondersteld dat de meeste ouders hun kinderen het liefst aanmelden bij scholen met een hoog percentage geslaagden. Vervolgens zouden deze scholen hun reputatie met gemak in stand kunnen houden door potentieel minder goede leerlingen te weigeren of tussentijds te laten vertrekken. Het gevaar voor dergelijk mechanismen wordt echter verkleind door de gegevens op een verantwoorde manier weer te geven. Vandaar dat Trouw het nodig vond om niet alleen de slagingspercentages maar ook andere cijfers in de krant te zetten.

Op grond van deze gegevens mag de Stedelijke Scholengemeenschap Roermond zich sinds kort de beste school van Nederland noemen. De school kreeg een negen, een negen en een acht voor respectievelijk de VWO-, HAVO- en MAVO-afdeling. Op bijvoorbeeld het VWO slaagde 97 procent van de leerlingen zonder in de vijfde klas te zijn blijven zitten. Het gemiddelde eindexamencijfer was een 7,0 en het percentage uitvallers en zittenblijvers 1 procent. Deze fraaie cijfers behaalde de school uit Roermond met 11 procent allochtone leerlingen. Verscheidene scholen kregen daarentegen een zware onvoldoende als rapportcijfer van Trouw. Op basis van deze cijfers kregen scholen in grote steden als Rotterdam en Utrecht, maar ook in kleine plaatsen als Lichtenvoorde en Sint Michielsgestel het predikaat 'slecht' opgeprikt.

De weergave van deze cijfers omlijstte Trouw met artikelen waarvan de koppen niets aan duidelijkheid overlieten, zoals 'Montessorischolen scoren slecht' of 'Mythen doorgeprikt' (Agerbeek et al., 1997, p. 1): "Over de kwaliteit van scholen doen vele indianenverhalen de ronde: veel allochtonen op school verminderen de prestaties, categorale gymnasia zijn beter dan (...) scholengemeenschappen en openbare scholen zijn slecht. In

wetenschappelijk onderzoek zijn die veronderstellingen vaak getoetst op basis van steekproeven, met soms wisselende conclusies. Nu alle scholen in de berekeningen kunnen worden meegenomen, blijken heel wat mythen doorgeprikt te kunnen worden."

In de onderstaande tekst vatten we de discussie over het Trouw-onderzoek samen. We geven de voornaamste voors en tegens in het debat weer. Vervolgens bespreken we de methode die Dronkers heeft gehanteerd om voor alle scholen een rapportcijfer te berekenen. We zullen daarbij enerzijds aandacht schenken aan de gekozen analysemethode en anderzijds ingaan op de gebruikte correctie voor verschillen in de leerlingenpopulatie van scholen. In een ander artikel (Dijkstra, Veenstra, Snijders & Peschar, 1998) laten we zien dat de keuze van een analysemethode gevolgen heeft voor de rapportcijfers van scholen. De opzet en de uitkomsten van dat artikel geven we hier beknopt weer. Op basis van die uitkomsten komen we tot een waardering van de uitvoering en conclusies van het Trouw-onderzoek. Onze probleemstelling verwoorden we als volgt: *Wat zijn de voor- en nadelen van het Trouw-onderzoek?*

2 Reacties op het Trouw-onderzoek: een overzicht

Op het Trouw-onderzoek kwamen positieve en negatieve reacties. In de eerste weken na de publicatie stonden er 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 uitkomsten 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 het onderzoek. Op dat commentaar gaan we nu uitgebreid in. Tegenstanders van het onderzoek menen dat de berichtgeving van het dagblad *tendentieus* is. Vooral voorstanders van Montessori-onderwijs vinden dat Trouw deze onderwijsvorm onrecht heeft aangedaan (Van der Hilst, 1997).

Verscheidene schoolleiders geven verder aan dat de cijfers geen rekening houden met de *bijzondere omstandigheden* waarin sommige scholen verkeren. Zo blijkt dat het Trouw-onderzoek geen rekening houdt 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) een verklaring voor het lage slagingspercentage van zijn VBO-afdeling met veel leerlingen die de individuele route volgen: "We begeleiden ze naar een baan, maar een VBO-diploma hebben ze dan niet op zak. Dat drukt het gemiddelde slagingspercentage enorm. 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."

Ook wordt er geen rekening gehouden met het sociaal milieu of het instroomniveau van leerlingen. Dergelijke gegevens heeft de inspectie namelijk niet. Door rekening te houden met de opleiding van ouders zullen de rapportcijfers volgens Dronkers echter niet met grote sprongen veranderen (Agerbeek et al., 1997, p. 1): "Het percentage allochtonen hangt sterk samen met het percentage laag geschoolde ouders dat kinderen op de school heeft. Mijn rapportcijfer zit heel dicht bij het ideaal." We zullen hier in de volgende paragraaf op terugkomen.

In andere reacties kwam naar voren dat het rapportcijfer *niet inzichtelijk* was. Van der Hilst (1997) 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." Doordat de formule aanvankelijk ontbrak, was het volstrekt onduidelijk waar de score van scholen op was gebaseerd. Een bestudering van de lijst leidde dan ook tot raadsels (vergelijk Pijls, 1998b). Zo kreeg bijvoorbeeld de VWO-afdeling van het Sint Jans College uit Heerlen een zes en die van het Fivel College uit Delfzijl een zeven. Op beide scholen is 2 procent van de leerlingen allochtoon, bedraagt het gemiddelde examen cijfer voor VWO een 6,8 en is het aantal uitvallers en zittenblijvers 3 procent. Het enige verschil tussen de twee scholen is het percentage onvertraagd geslaagden. Op het Fivel College, de school met een zeven, slaagt 82 procent van de leerlingen onvertraagd en op het Sint Jans College, de school met een zes, 92 procent van de leerlingen! De oorzaak hiervan is dat in het rapportcijfer verschillende indicatoren voor schoolsucces worden samengenomen met een berekeningsmethode waarbij kanttekeningen en vraagtekens kunnen worden geplaatst. Pijls (1998a) vraagt zich bijvoorbeeld af waarom in de berekening van Dronkers Nederlands op het VBO 57 keer zo zwaar telt als Engels of waarom meer tussentijdse uitvallers in VWO-3 een hoger rapportcijfer oplevert. De psycholoog Vervaeke noemt het Trouw-onderzoek onwetenschappelijk (Van Calmhout, 1997): "Er wordt een

indrukwekkende formule met weet ik hoeveel parameters gebruikt waarvan niemand weet of die aantoonbaar enig hout snijdt. Waarom is het die formule en niet een heel andere?"

Ook bleek dat sommige gegevens over scholen *foutief of twijfelachtig* van kwaliteit zijn. Zo is het percentage allochtone leerlingen op discutabele wijze tot stand gekomen (Louwerse, 1997). Een eerste probleem met dit percentage is dat het niet per schooltype en ook niet per locatie, maar voor de gehele school geldt. Een tweede probleem is dat verscheidene leerlingen als autochtoon worden beschouwd, hoewel ze toch een andere etnische afkomst hebben. Dit geldt met name voor Surinaamse, Antilliaanse en Arubaanse leerlingen. Over de examencijfers voor VBO en MAVO kan worden opgemerkt dat er geen rekening wordt gehouden met het niveau waarop leerlingen het examen afleggen. Op sommige scholen legt bijvoorbeeld meer dan de helft van de MAVO-leerlingen alle vakken op het hoogste niveau af. Op andere scholen doet geen enkele leerling op dat niveau in alle vakken examen. De cijfers op de verschillende niveaus kunnen echter eenvoudig worden omgerekend tot een cijfer voor alle niveaus. Een 6 op D-niveau staat dan bijvoorbeeld ongeveer gelijk aan een 7 op C-niveau. Ten slotte zijn 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.

Om te kunnen spreken van kwaliteit van scholen, zo haalde ook menigeen aan, moet er op zijn minst sprake zijn van cijfers over verscheidene jaren. 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.

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. In de publicatie van het Trouw-onderzoek ontzenuwt de redactie dergelijke bezwaren al bij voorbaat (Agerbeek et al., 1997, p. 6): "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 schoolprestatie lijst is dan ook een steuntje in de rug bij de schoolkeuze, niets meer en niets minder."

Het rapportcijfer is dan ook niet het waardeoordeel over bijvoorbeeld het didactisch handelen van een school, maar een cijfer gebaseerd op schoolprestaties. Een ingezonden reactie naar Trouw (Giesen, 1997) geeft daarvan een illustratie. Zo blijkt dat het Marnix College in Ede twee keer voorkomt in het overzicht voor VWO-scholen. De school is namelijk een gemeenschap van scholen, te weten een gymnasium en een school voor VWO, HAVO en MAVO. De twee VWO-afdelingen van het Marnix College krijgen twee sterk verschillende rapportcijfers: een acht en een vier. Deze verschillende waarderingen zeggen alleen maar iets over een toevallige groep eindexamenleerlingen, want de leerlingen van de twee afdelingen zitten bij elkaar in de klas. Op de vakken Grieks en Latijn na volgen de VWO-leerlingen dezelfde lessen, krijgen ze dezelfde begeleiding en worden ze op dezelfde manier gecorrigeerd.

De suggestie moet ook niet zijn dat verschillen in prestaties van leerlingen vooral op het conto van een school moeten worden geschreven. Uit onderzoek blijkt dat die verschillen slechts in geringe mate samenhangen met verschillen tussen scholen (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld & York, 1966; Luyten, 1994). Uit de resultaten van onderzoek blijkt dat na correctie voor instroomkenmerken de tussenschoolse variantie 10 tot 15 procent bedraagt en de binnenschoolse variantie 85 tot 90 procent. Dit betekent 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.

Enkele scholen waren zo verontwaardigd over het verkregen rapportcijfer dat ze van Trouw een rectificatie eisten. Het Sittards College scoorde voor de HAVO-, de atheneum- en de gymnasium-afdeling respectievelijk een zes, een vier en een zeven. In het Dagblad De Limburger (Redactie, 1997) stelt de directie dat de school verkeerd is beoordeeld. "(S)cholieren 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." De

schoolleiding vindt dan ook dat de goede naam van de school ten onrechte is aangetast. Het Da Vinci College uit Leiden dreigde zelfs naar de rechter te stappen om door een kort geding rectificatie af te dwingen (Koevoet, 1997). De school kreeg een zeven voor de VWO-afdeling en een drie voor de HAVO-afdeling. De onvoldoende van de HAVO was het gevolg van een onwaarschijnlijk hoog percentage zittenblijvers en uitvallers. Vooral vanwege de opvang van asielzoekers is per leerjaar 17 procent van de HAVO-leerlingen zittenblijver of uitvaller. Vier weken na de publicatie van het Trouw-onderzoek stelde Dronkers opnieuw een rapportcijfer voor de HAVO-afdeling van het Da Vinci College vast. In zijn nieuwe berekening telden de asielzoekers niet mee en kwam hij uit op een acht.

In een reactie op het Trouw-onderzoek heeft het ministerie van Onderwijs beloofd om in 1998 informatie uit te brengen over de kwaliteit van scholen voor voortgezet onderwijs (Netelenbos, 1997). Scholen zullen, zoals al geruime tijd de bedoeling was, schoolgidsen gaan uitgeven met daarin gegevens over bijvoorbeeld de schoolprestaties, de aandacht voor leerlingbegeleiding en de buitenschoolse activiteiten. In aanvulling daarop zal het ministerie zogeheten kwaliteitskaarten gaan uitbrengen. De kwaliteitskaart zal per school informatie geven over de resultaten, de organisatie, het aanbod, het onderwijsleerproces en de omgeving. De gegevens waarmee de inspectie de kwaliteitskaarten samenstelt, worden vooraf aan de scholen ter controle voorgelegd, zodat eventuele fouten kunnen worden gecorrigeerd. Behalve kwaliteitskaarten wil het ministerie vanaf dit jaar ook regionale en landelijke overzichten publiceren met de resultaten van scholen.

In deze kwaliteitsmeting vindt Dronkers (1997a) twee punten volstrekt onvoldoende, namelijk het ontbreken van zowel een samenvattend oordeel als een systematische correctie voor verschillen in leerlingenpopulatie. Op dit laatste punt gaan we in het onderstaande in als we het hebben over het belang van netto indicatoren van schoolprestaties. Het ontbreken van een samenvattend oordeel, zoals het rapportcijfer van het Trouw-onderzoek, vindt Dronkers funest voor ouders. "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." Volgens Dronkers (1997a) is het voorstel uitsluitend gericht op de belangen van scholen en niet op die van ouders.

3 Meten van schoolprestaties

3.1 Het rapportcijfer van het Trouw-onderzoek

Twee weken na de publicatie van het Trouw-onderzoek publiceerde Dronkers de details van de berekeningsmethode waarmee het rapportcijfer tot stand kwam¹. Dronkers (1997b) toont dat het rapportcijfer een cijfer is voor het totale rendement van een school, of eventueel van een schooltype binnen een school. Dit rendement is een optelsom van examencijfers, het slagingspercentage, het uitvalpercentage (twee typen) en het percentage zittenblijvers. Voor VWO-, HAVO-, MAVO- en VBO-afdelingen hanteert hij respectievelijk vijftien, dertien, tien en negen kenmerken.

Per schooltype stelt hij een rapportcijfer samen, waarbij geldt dat het rapportcijfer hoger is naarmate de eindexamencijfers hoger zijn, het slagingspercentage hoger is en de percentages zittenblijvers en uitvallers lager zijn. Het rapportcijfer is daarbij gecorrigeerd voor het percentage allochtone leerlingen op een school. Voor bijvoorbeeld het VBO neemt het rapportcijfer met één punt toe bij 25 procent meer allochtone leerlingen. Bij 12 procent meer allochtone leerlingen neemt het VWO-rapportcijfer met dezelfde hoeveelheid toe.

Voor de berekening van het rapportcijfer heeft Dronkers onder meer factor analyse en regressieanalyse gebruikt. Uit de uiteindelijke analyses komt per school een score. Bij die score telt 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) 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 cijfers². Wanneer er binnen een bepaald schooltype minder dan vijftig eindexamenkandidaten zijn, heeft hij die school geen rapportcijfer gegeven.

3.2 Een replicatie van het Trouw-onderzoek

Om na te gaan in hoeverre de rapportcijfers een betrouwbaar beeld geven van de prestatieverschillen tussen scholen, hebben we een tweevoudige strategie gevolgd, waarvan in een ander artikel verslag wordt gedaan (Dijkstra et al., 1998). In eerste instantie hebben we analoog aan Dronkers een factor- en regressieanalyse uitgevoerd en zijn we nagegaan wat de effecten zijn van een meer en minder uitvoerige controle voor de kenmerken van de leerlingenpopulatie. In tweede instantie hebben we gebruik gemaakt van een alternatieve analysemethode, waarbij we de rapportcijfers voor de scholen met

multiniveau analyse hebben berekend. Ook bij deze methode zijn we nagegaan wat de gevolgen zijn van een meer of minder uitgebreide correctie voor instroomverschillen tussen scholen. De analyse bestaat dus uit vier stappen:

1A: methode Trouw - correctiefactoren Trouw

1B: methode Trouw - alternatieve correcties

2A: alternatieve methode - correctiefactoren Trouw

2B: alternatieve methode - alternatieve correcties

In onze replicatie (Dijkstra et al., 1998) hebben we geen gebruik gemaakt van de inspectie-gegevens waarop het onderzoek van Trouw is gebaseerd. Een belangrijke beperking daarvan is immers de beperkte beschikbaarheid van informatie over de achtergrondkenmerken van leerlingen. Bij het berekenen van een rapportcijfer kon Dronkers namelijk uitsluitend voor het percentage allochtone leerlingen corrigeren. We hebben daarom een recente dataset, het cohortonderzoek VOCL'89, gebruikt met uitgebreide informatie over leerlingen en hun schoolloopbaan in het voortgezet onderwijs (Driessen & Van der Werf, 1992). Hierbij hebben we gegevens van 19.260 leerlingen van 381 scholen gebruikt.

Enkele jaren geleden heeft bijvoorbeeld Bosker (1990, 1992) al eens laten zien dat een beoordeling van de effectiviteit van scholen niet op *bruto* maar op *netto indicatoren* van schoolprestaties gebaseerd dient te zijn. Om het effect van een school op haar leerlingen te achterhalen, moet rekening worden gehouden met de achtergrondkenmerken van de leerlingenpopulatie. Bij de cijfers in Trouw over het percentage onvertraagd geslaagden, het aantal uitvallers en zittenblijvers en de gemiddelde eindexamencijfers is dat niet gebeurd. Strikt genomen laten dergelijke bruto indicatoren dus geen uitspraken toe over de kwaliteit van een school. We hebben onze replicatie daarom alleen gericht op het door *Trouw* uitgedeelde rapportcijfer. Met onze analyses willen we nagaan in hoeverre een alternatieve aanpak van de rapportcijfers tot verschuivingen in de beoordeling van scholen leidt.

Voor de stappen 1A en 1B hebben we vier maten voor schoolprestaties gehanteerd, te weten of een leerling onvertraagd het examen heeft gehaald, het aantal keren dat een leerling is blijven zitten, of een leerling al dan niet voortijdig de school heeft verlaten en de positie van een leerling op een zogeheten leerjarenladder. Deze maten hebben we berekend per schooltype, waarbij we als richtlijn hebben gehanteerd dat er minimaal tien leerlingen binnen één schooltype vallen.

De eerste drie prestatiematen zijn ook in het Trouw-onderzoek gebruikt. De *leerjarenladder* (Bosker, Van der Velden & Hofman, 1985) is een extra maat. Met deze leerjarenladder kunnen we nagaan welk niveau leerlingen sinds het schooljaar 1989-'90 hebben

bereikt. De leerjarenladder drukt de afstand in leerjaren uit tot de top van het onderwijssysteem. Het maximale niveau is negen, wat kan worden bereikt door leerlingen die zonder doubleren het VWO hebben doorlopen. Bij de opstelling van de leerjarenladder hebben Bosker et al. rekening gehouden met zittenblijven en doorstroom. Een leerling die blijft zitten verliest een jaar op de overigen. Bij doorstroming van bijvoorbeeld MAVO-4 naar HAVO-4 doet een leerling een jaar langer over het bereiken van dat niveau dan leerlingen die de HAVO regulier doorlopen.

In de stappen 1A en 1B hebben we de methode van het Trouw-onderzoek gevolgd. Met factoranalyse hebben we één succesmaat per school geconstrueerd uit de verschillende prestatie-maten. In stap 1A hebben we met regressieanalyse deze succesmaat gecorrigeerd voor een relevant kenmerk van de leerlingenpopulatie, namelijk het percentage allochtone leerlingen. Vervolgens hebben we in stap 1B ook rekening gehouden met de gemiddelde prestaties bij instroom en het gemiddelde opleidingsniveau van de vader.

Uit de uitkomsten van de stappen 1A en 1B blijkt *voorlopig* dat er geringe verschillen zijn in rapportcijfer tussen een meer of minder uitgebreide correctie voor populatieverschillen. Kleine verschillen tussen de twee stappen zijn er alleen voor VBO- en HAVO-scholen. In Tabel 1 maken we een vergelijking tussen de stappen 1A en 1B. Op schoolniveau blijkt de samenhang tussen het percentage allochtone leerlingen, de gemiddelde sociaal-economische status en het gemiddelde prestatieniveau bij intrede zo hoog dat extra correcties de rapportcijfers nauwelijks veranderen, zie Bijlage A.

(HIER ONGEVEER TABEL 1 PLAATSEN)

Voor de stappen 2A en 2B hebben we de leerjarenladder als succesmaat gehanteerd. In een factoranalyse van de vier prestatie-maten worden namelijk variabelen samengenomen die niet één dimensie vormen. Ze bevatten elementen van zowel echte prestatie als sturingsbeleid. De mate waarin scholen een streng beleid voor de overgang voeren is op zich niet relevant voor een prestatie-maat. Een scholier die zeven jaar over het VWO doet heeft deze verblijfsduur op school ongeacht of hij een keer is blijven zitten of een keer voor zijn eindexamen is gezakt. In de factoranalyse wordt succes dus gecontamineerd met het door de school gehanteerde sturingsmechanisme.

Ook hebben we voor een andere analyse-methode gekozen, namelijk multiniveau analyse. In een dergelijke analyse blijven het leerlingniveau en het schoolniveau beide in onderlinge samenhang aanwezig en kunnen de variantiecomponenten per niveau worden geschat (Bryk & Raudenbush, 1992; Van den Eeden & Meijnen,

1990). Bosker (1990, p. 38 e.v.) geeft een inzichtelijke illustratie van het gevaar van geaggregeerde data, hetgeen de gegevens van het *Trouw*-onderzoek zijn. Hij laat zien hoe gebruik van gegevens op het school- dan wel leerlingniveau tot tegengestelde conclusies over de effectiviteit van scholen kan leiden.

Het onderscheid tussen de analyse van geaggregeerde data en multiniveau analyse is op twee punten relevant, te weten de controle voor de samenstelling van de leerlingenpopulatie en de nauwkeurigheid van de resultaten (Dijkstra et al., 1998). Om na te gaan in hoeverre gebruik van deze in statistisch opzicht meer geëigende methode effect heeft op de succesmaat van scholen, hebben we de stappen 2A en 2B gebaseerd op een multiniveau analyse, uitgevoerd met het programma HLM (Bryk, Raudenbush & Congdon, 1993). Hierbij hebben we gebruik gemaakt van de naar z-scores omgezette residuen die we met HLM voor het intercept op schoolniveau hebben berekend.

Analoog aan de werkwijze van Dronkers hebben we in stap 2A alleen gecorrigeerd voor het percentage allochtone leerlingen van de school. In stap 2B hebben we extra correcties toegevoegd: de variabelen etniciteit, opleiding van de vader en de aanvankelijke leerprestaties op leerlingniveau; de variabelen percentage allochtone leerlingen, gemiddeld opleidingsniveau van de vader en gemiddelde prestaties bij instroom op schoolniveau.

De verschillen tussen de in stap 2A en 2B berekende scores op de succesmaat geven we weer in Tabel 1. De meer uitgebreide controle in stap 2B leidt tot aanzienlijke verschillen in de scores van scholen. Voor veel scholen treden verschillen van één of meer punten op. De verschillen zijn het grootst voor MAVO- en HAVO-scholen.

Vergelijking van beide methoden (*1A-2A*, *1B-2B*) in Tabel 1 laat zien dat gebruik van een multiniveau analyse met de leerjarenladder als succesmaat, waarvoor we pleiten, tot een aanzienlijke verschuiving in de scores van scholen leidt. De variabelenvergelijking maakt duidelijk dat effecten van verschillen in correctie voor een meer of minder kansrijke leerlingenpopulatie zich vooral voordoen bij de op multiniveau analyse gebaseerde succesmaten (*2A-2B*). Dergelijke verschillen treden bij een op schoolniveau geaggregeerde regressieanalyse nauwelijks op (*1A-1B*).

(HIER ONGEVEER TABEL 2 EN 3 PLAATSEN)

In Tabel 2 en 3 vergelijken we per schooltype de rapportcijfers van enerzijds de op schoolniveau geaggregeerde regressieanalyse met beperkte correctie (stap 1A) en anderzijds de multiniveau analyse

met uitgebreide correctie (stap 2B). Voor het VWO geldt bijvoorbeeld dat 43 procent van de scholen in beide stappen hetzelfde cijfer heeft. Deze scholen zijn weergegeven op de diagonaal in Tabel 2. De overige VWO-scholen hebben verschillende cijfers in beide stappen, waarbij 26 procent van de scholen een onvoldoende behaalt in de stap die analoog is aan het Trouw-onderzoek en een voldoende scoort in de door ons bepleite stap. Omgekeerd scoort 6 procent van de VWO-scholen in stap 1A voldoende en in stap 2B onvoldoende.

Tussen de vier succesmaten bestaan dus opmerkelijke verschillen. Een voorzichtige interpretatie van rapportcijfers gebaseerd op een berekeningswijze zoals in het dagblad *Trouw* is dan ook geboden. De hier gepresenteerde analyses laten zien dat voor veel scholen aanpassing van de succesmaten met één of twee punten allerminst denkbeeldig is.

3.3 Een voortbestaan van mythen?

Wanneer we vaststellen dat de methode voor berekening van een rapportcijfer in het Trouw-onderzoek niet zonder beperkingen is, willen we ook aan de conclusies die daarop zijn gebaseerd niet voorbijgaan.

Over bijvoorbeeld de richting van scholen doet de redactie uitspraken. Gereformeerd-vrijgemaakte en reformatorische scholen blijken in het Trouw-onderzoek bovengemiddeld te scoren. Daarentegen scoren Montessori- en algemeen bijzondere scholen volgens Trouw onder het gemiddelde. De overige schooltypen scoren ongeveer op gemiddeld niveau. Sinds de jaren tachtig wordt er veelvuldig onderzoek gedaan naar de verschillen tussen openbaar en bijzonder onderwijs. In de Verenigde Staten concluderen onderzoekers (Coleman & Hoffer, 1987; Bryk, Lee & Holland, 1993) dat katholieke scholen effectiever zijn dan andere scholen in de particuliere sector of scholen in de publieke sector. Dit effect kan gedeeltelijk worden verklaart door verschillen in school- en klassenkenmerken tussen katholieke en andere scholen. Dergelijke kenmerken gebruiken deze onderzoekers dan ook om het succes van katholieke scholen te verklaren. In Nederland blijkt het minder gemakkelijk om een algemene uitspraak te doen over het succes van een bepaalde richting in het voortgezet onderwijs (Dijkstra, 1997). In de discussie over het mogelijke succes van het bijzondere onderwijs en het mogelijke falen van het openbare onderwijs is er dan ook allerminst sprake van overeenstemming.

Gebaseerd op empirisch onderzoek, zouden zo bij meer conclusies uit het Trouw-onderzoek kanttekeningen te plaatsen zijn. We moeten dan ook lachen om Wansink (1997) als hij de mislukking van sociaal-wetenschappelijk onderzoek beschrijft. "(D)e onderwijswetenschap wordt door Trouw te kijk gezet. Bij gebrek aan

exacte gegevens, was het gemakkelijk promoveren in deze tak van sport. Generalisaties over witte en zwarte scholen, over openbare en bijzondere scholen, over kindgericht en klassikaal lesgeven, en over scholengemeenschappen en zelfstandige gymnasia konden welig tieren."

Tot slot willen we naar voren brengen dat op basis van geaggregeerde gegevens geen individuele uitspraken kunnen worden gedaan. Op basis van het Trouw-onderzoek wordt gesteld dat brede scholen goed uitpakken voor HAVO en VWO, maar de vraag moet natuurlijk zijn of brede scholen ook goed uitpakken voor potentiële HAVO- en VWO-leerlingen. Als een leerling in zeven jaar tijd een VWO-diploma behaalt, dan doet zijn school het volgens het Trouw-onderzoek slechter dan wanneer die leerling in vier jaar voor zijn MAVO-examen slaagt. Prick (1997) zegt hierover: "De school die streng selecteert, die vindt dat alleen een smalle top in staat is een VWO-diploma te halen, krijgt hoge cijfers; de school die ook minder getalenteerden, wellicht moeizamer, maar toch aan zo'n diploma helpt wordt gediskwalificeerd. Dat is natuurlijk de omgekeerde wereld."

Met een multiniveau analyse heeft Kreft (1993) gekeken naar de onderwijskansen van leerlingen in verschillende vormen van voortgezet onderwijs. Uit haar analyses blijkt dat de Cito-score een sterk voorspellende waarde heeft voor de onderwijskansen van leerlingen in heterogene brugklassen. Naarmate leerlingen een hogere score op die toets hebben, komen ze uiteindelijk in hogere vormen van voortgezet onderwijs terecht. In homogene schooltypen, zoals een zelfstandig gymnasium of een categorale MAVO, voorspelt de Cito-score de onderwijskansen van leerlingen minder goed. Op een gymnasium hebben leerlingen met een lage Cito-score meer kans om een VWO-diploma te halen dan op een brede scholengemeenschap. Daarentegen komen leerlingen met een hoge Cito-score minder ver op een categorale MAVO dan op een brede scholengemeenschap. Waar Trouw concludeert dat brede scholen goed uitpakken voor HAVO en VWO, stelt Kreft dat brede scholen minder gunstig zijn voor potentiële HAVO- en VWO-leerlingen, dat wil zeggen leerlingen met een hoge Cito-score.

4 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. De door het dagblad gepubliceerde hitlijst trok veel aandacht en leidde tot een uitgebreide discussie in de media. Zo kreeg de Trouw-redactie lof, omdat de resultaten van scholen nu

eindelijk openbaar werden en een belangrijke stap naar publieke verantwoording van de kwaliteit van scholen was gezet. Ook constateerde menigeen met vreugde dat niet alleen slagingspercentages maar ook andere uitkomsten van scholen in de beoordeling betrokken waren. Daarnaast vielen echter nogal wat negatieve reacties te beluisteren. Zo werd Trouw tendentieuze berichtgeving verweten, zouden de gebruikte gegevens nogal wat onjuistheden bevatten en ontstond een discussie over schoolprestaties als een te beperkte interpretatie van wat goede scholen hun leerlingen bieden. Andere commentaren richtten zich op het rapportcijfer dat aan scholen was toegekend en de geringe inzichtelijkheid daarvan, en de vraag of in de beoordeling van scholen wel voldoende rekening wordt gehouden met de soms aanzienlijke verschillen die tussen leerlingenpopulaties bestaan.

Het blijkt echter dat juist 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 multiniveau analyse zo maar twee punten lager of hoger scoren. In de verdere ontwikkeling van methoden om de kwaliteit van scholen inzichtelijk te maken, pleiten wij voor een multiniveau analyse, omdat alleen met een dergelijke methode adequaat kan worden gecontroleerd voor de samenstelling van de leerlingenpopulatie. Ook zijn de resultaten van een multiniveau analyse nauwkeuriger dan de uitkomsten van een regressieanalyse.

Hierbij verdient het de voorkeur om een leerjarenladder of een soortgelijke maat als indicator van de prestaties van scholen te gebruiken. Een dergelijke maat geeft alleen het succes van scholen weer. In het Trouw-onderzoek zijn variabelen samengenomen die enerzijds het succes en anderzijds de sturingsmechanismen van scholen weergeven. Wij vinden dan ook dat deze variabelen niet één dimensie vormen. Door de lage hoeveelheid verklaarde variantie van de eerste factor wordt deze conclusie ondersteund.

Wanneer we kiezen voor een multiniveau analyse dan blijkt dat de correctie voor alleen het percentage allochtone leerlingen mager is. Het rapportcijfer dient te worden verfijnd met meer gegevens over de kwaliteit van de instroom van scholen. Bosker (1992) heeft deze aanbeveling al eerder gedaan en in de publicatie van het Trouw-onderzoek (Agerbeek et al., 1997) wordt er ook gewag van gemaakt. Wij willen er echter op wijzen dat een uitgebreide controle met name zin heeft in een multiniveau analyse.

Wanneer Trouw of de overheid opnieuw een onderzoek gaat doen naar de prestaties van scholen met gebruik van dezelfde methode, dan kan men zich eigenlijk wel de kosten en de moeite besparen om extra gegevens te verzamelen, want met succesmaten

die zijn gebaseerd op een op schoolniveau geaggregeerde regressieanalyse maakt het nauwelijks uit of je meer of minder uitgebreid controleert. De keuze voor die methode verwerpen wij.

Met het publiceren van gegevens over schoolprestaties heeft Trouw een doorbraak geforceerd in het publieke debat over de kwaliteit van scholen. Het Trouw-onderzoek heeft het onderwerp schoolprestaties weliswaar in het middelpunt van de belangstelling geplaatst, maar het heeft niet tot nieuwe inzichten geleid. Er zijn geen mythen doorgeprikt, sterker nog er zijn mogelijk nieuwe mythen geschapen. Ineens kwamen er verhalen in de wereld waaruit bleek dat Montessori-scholen het slecht zouden doen of dat brede scholengemeenschappen goed zouden zijn voor HAVO en VWO. Dergelijke conclusies mogen echter helemaal niet worden getrokken uit een onderzoek waar het mankeert aan een correcte succesmaat en analysemethode. Met multiniveau analyse en een leerjarenladder zal in toekomstig onderzoek hopelijk opnieuw worden gekeken naar bijvoorbeeld de verschillen tussen Montessori-scholen en andere scholen.

We hebben ons uiteraard afgevraagd of het Trouw-onderzoek maatschappelijk en wetenschappelijk verantwoord was. Op basis van dit artikel stellen we dat de wetenschappelijke fundamenten van het Trouw-onderzoek niet sterk zijn. We stemmen dan ook in met hetgeen Dronkers op de Twaalfde Onderwijs sociologische Conferentie, gehouden op 19 en 20 november 1997, naar voren bracht. Tijdens een debat over het Trouw-onderzoek maakte hij duidelijk dat zijn bijdrage aan het onderzoek niet moest worden opgevat als sociaal-wetenschappelijk onderzoek, maar als maatschappelijke dienstverlening met behulp van wetenschappelijke inzichten. Wij wijzen er echter op dat dit in het Trouw-onderzoek niet duidelijk naar voren kwam.

Maatschappelijk gezien kleven er ook bezwaren aan het Trouw-onderzoek. We vinden het onjuist dat de scholen voor voortgezet onderwijs niet betrokken zijn bij de publicatie. In een consumentenonderzoek is het een goed gebruik om producenten te informeren over de cijfers die hun producten krijgen toebedeeld. Eventuele onjuistheden in het onderzoek kunnen dan voor de publicatie worden bijgesteld en de juiste gegevens kunnen dan met elkaar worden vergeleken. Bij het opstellen van de kwaliteitskaarten zal een dergelijke procedure wel worden gevolgd. Voor consumenten is niet alleen de openbaarheid, maar ook de correctheid van dergelijke gegevens van belang. De gegevens van de inspectie bleken echter in talrijke gevallen foutief of twijfelachtig te zijn.

Ook was het goed geweest wanneer er in overleg met de scholen was nagedacht over hoe je goede en slechte resultaten publiek maakt. Het is daarom nodig ze te betrekken bij het onderzoek, zodat

ze in een vroeg stadium weten wat het doel, de opzet en de wetenschappelijke basis van het onderzoek is en hoe de rapportage van het onderzoek eruit gaat zien. Eventueel commentaar op het onderzoek hadden schoolleiders dan voor de publicatie kunnen geven.

Een dergelijke wijze kost natuurlijk tijd. Voor een rapportage over de schoolprestaties in verschillende landen heeft de *Organisation for Economic Cooperation and Development* (OECD) daar de tijd voor genomen. Zoals scholen het niet leuk vinden om een onvoldoende te krijgen, wil geen land onderaan komen op een internationale rangorde. De OECD heeft daar echter rekening mee gehouden en er vier jaar over gedaan om een internationale vergelijking te maken. Een dergelijk traject heeft Trouw, naar wij aannemen, nooit voor ogen gehad. Uit journalistiek oogpunt valt dat te begrijpen, maar of het maatschappelijk verstandig was betwijfelen we.

Of we hiermee aan het begin van een nieuw tijdperk staan, waarin steeds meer ouders als onderwijsconsument gaan optreden is twijfelachtig. Helpt het Trouw-onderzoek of een verbeterde versie daarvan ouders bij hun keuze voor een school? Moeten scholen die hoge cijfers hebben behaald rekenen op een grotere toestroom van leerlingen? Zijn scholen die als slecht naar voren zijn gekomen in het komende schooljaar aangewezen op het ontslaan van verscheidene docenten? Agerbeek (1997b), redacteur van Trouw, denkt dat het niet zo'n vaart zal lopen. "De angst dat ouders hun kinderen niet naar scholen met een slecht rapportcijfer zullen sturen, zou wel eens ongegrond kunnen blijken. Ze weten best dat er op een school meer te koop is dan een rapportcijfer in een krant." Sikkes en Westenberg (1997) stellen dat ouders eerder van een keuze afzien vanwege bepaalde incidenten. "Echt grote gevolgen voor de school heeft de publicitaire ramp die ontstaat na incidenten rondom de school. Een steekpartij op de rand van het schoolplein. Leerlingen die opgepakt worden voor drugshandel. De rector die tijdens een eindexamenstunt de bh van de stripteasedanseres losknoopt en zelf ook in z'n onderbroek op het podium eindigt. Dat straffen ouders af. Dat kan tientallen leerlingen kosten."

Het is de vraag of scholen hun beleid gaan richten naar de criteria van dit soort onderzoek. Scholen zouden namelijk op de cijfers kunnen anticiperen door in geval van twijfel leerlingen te verwijzen naar lagere vormen voor voortgezet onderwijs. Voor het succes van een school is het immers beter dat een leerling in vier jaar de MAVO haalt, dan dat een leerling na zes jaar een HAVO-diploma behaalt. Voorlopig is er geen reden om te denken dat scholen hun cijfer zelf proberen te bepalen. Wanneer echter zou blijken dat de cijfers het keuzegedrag van ouders sterk beïnvloeden dan zouden scholen wel eens strategisch te werk kunnen gaan.

In de inleiding gaven we aan dat de ongelijkheid tussen scholen kan worden vergroot door de hitlijsten. In Nederland zou het ook zo kunnen gaan. Dergelijke onbedoelde effecten moeten worden meegenomen in een oordeel over het Trouw-onderzoek en de toekomstige kwaliteitskaarten. Het zou toch tragisch zijn wanneer de publicatie van schoolresultaten zou leiden tot een toename van ongelijkheid in de samenleving?

Noten

1. De opzet van de berekening was overigens van begin af aan op te vragen bij Dronkers (Agerbeek, 1997a).
2. 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 vinden wij niet gelukkig, omdat kleine verschillen tussen scholen dan groter kunnen lijken dan ze werkelijk zijn.

Literatuur

- Agerbeek, M. (1997a). De meest gestelde vragen over het *Trouw*-onderzoek 'Schoolprestaties'. *Trouw*, 29 oktober, p. 17.
- Agerbeek, M. (1997b). *Trouw*-onderzoek Schoolprestaties: antwoord op de emoties. *Trouw*, 8 november, p. 12.
- Agerbeek, M., Hageman, E., Kreulen, E., & Lakmaker, H. (1997). *Trouw*-onderzoek schoolprestaties. *Trouw*, 25 oktober, ZENZ.
- Bosker, R.J. (1990). Extra kansen dankzij de school? Nijmegen: ITS (academisch proefschrift Rijksuniversiteit Groningen).
- Bosker, R.J. (1992). Outputcontrole en kansenongelijkheid. In: Ministerie van Onderwijs en Wetenschappen, Ceders in de tuin: Voorstudies: naar een nieuwe opzet van het onderwijsbeleid voor allochtone leerlingen (hoofdstuk vijf). 's-Gravenhage: DOP.
- Bosker, R.J., Velden, R.K.W. van der, & Hofman, W.H.A. (1985). Een generatie geselecteerd. Deel 1: De loopbanen. Groningen: RION.
- Bryk, A.S., Lee, V.E., & Holland, P.B. (1993). Catholic schools and the common good. Cambridge, MA: Harvard University Press.
- Bryk, A.S., & Raudenbush, S.W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

- Bryk, A.S., Raudenbush, S.W., & Congdon, R.T. (1993). HLM2 and HLM3. Computer programs and users' guide. Chicago: University of Chicago.
- Calmhout, M. van (1997). Plaaggeesten bestaan. de Volkskrant, 8 november, p. 1W.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). Equality of educational opportunity. Washington DC: US Government Printing Office.
- Coleman, J.S., & Hoffer, T. (1987). Public and private high schools: The impact of communities. New York: Basic Books.
- Dijkstra, A.B. (1996). "Markt of arena? Over onbedoelde effecten van grotere onderwijsvrijheid". In : S. Miedema, & H. Klifman (red.), Christelijk onderwijs in ontwikkeling (pp. 93-114). Kampen: Kok.
- Dijkstra, A.B. (1997). "Onderwijskansen en richting van de school". In: A.B. Dijkstra, J. Dronkers, & R. Hofman (red.), Verzuiling in het onderwijs: Actuele verklaringen en analyse (pp. 144-184). Groningen: Wolters-Noordhoff.
- Dijkstra, A.B., Veenstra, R., Snijders, T., & Peschar, J. (1998). Trouw moet blijken. Opmerkingen bij een hitlijst voor scholen. Tijdschrift voor Onderwijsresearch, 23, in druk.
- Driessen, G., & Werf, G. van der (1992). Het functioneren van het voortgezet onderwijs. Beschrijving steekproef en psychometrische kwaliteit instrumenten. Groningen/Nijmegen: RION/ITS.
- Dronkers, J. (1997a). Netelenbos komt met volstrekt onvoldoende kwaliteitskaart van scholen. Trouw, 31 december, p. 11.
- Dronkers, J. (1997b). Schoolcijfer zegt niet dat scholen met 5 onvoldoende functioneren. Trouw, 8 november, p. 12.
- Eeden, P. van den, & Meijnen, G.W. (red.) (1990). Multi-niveauonderzoek: uitgangspunten en toepassingen. Themanummer. Tijdschrift voor Onderwijsresearch, 15(5), 257-333.
- Giesen, H. (1997). Schoolprestaties (1), Trouw, 29 oktober, p. 11.
- Hilst, B. van der (1997). Montessori-scholen doen het beter dan hun omgeving, Trouw, 6 november, p. 11.
- Koevoet, W. (1997). Da Vinci College stapt naar rechter. Leidsch Dagblad, 30 oktober, p. 4.
- Kreft, I.G.G. (1993). Using multilevel analysis to assess school effectiveness: a study of Dutch secondary schools. Sociology of Education, 66, 104-129.
- Louwerse, P. (1997). De venijnige wraak op professor Dronkers. Rotterdams Dagblad, 10 december, p. 5.

- Luyten, H. (1994). School effects - stability and malleability. Enschede: s.n. (academisch proefschrift Universiteit Twente).
- Netelenbos, T. (1997). Kwaliteit. Uitleg, 15(26), 5 november, p. 15.
- Pijls, W. (1998a). Dronkers maakt kwaliteitskeurmerk scholen nog altijd niet duidelijk. Trouw, 3 januari, p. 11.
- Pijls, W. (1998b). Schoolcijfers van Dronkers slaan helemaal nergens op. de Volkskrant, 19 februari, p. 9.
- Prick, L. (1997). Vergelijking van scholen is krakkemikkige vertoning. NRC Handelsblad, 1 november, p. 7.
- Redactie (1997). Sittard College eist rectificatie Trouw. Dagblad De Limburger, 29 oktober, p. 7.
- Redactie onderwijs (1997). Kille cijfers kloppen maar komen hard aan. NRC Handelsblad, 27 oktober, p. 3.
- Sikkes, R., & Westenberg, J. (1997). Voldoende of onvoldoende. de Volkskrant, 8 november, p. 5V.
- Wansink, H. (1997). Scholen houden ramen liever dicht. de Volkskrant, 10 november, p. 7.

Auteurs

D.R. Veenstra is assistent-in-opleiding bij de Vakgroep Sociologie van de Rijksuniversiteit Groningen en maakt deel uit van de onderzoekschool ICS.

A.B. Dijkstra is onderzoeker bij de Vakgroep Sociologie van de Rijksuniversiteit Groningen en maakt deel uit van de onderzoekschool ICS.

J.L. Peschar is hoogleraar onderwijssociologie bij de Vakgroep Sociologie van de Rijksuniversiteit Groningen en maakt deel uit van de onderzoekschool ICS.

T.A.B. Snijders is hoogleraar stochastische modellen voor de gedrags- en maatschappijwetenschappen bij de Vakgroep Statistiek en Meettheorie van de Rijksuniversiteit Groningen en is bestuurslid van de onderzoekschool ICS.

Correspondentie naar de eerste auteur: ICS/Vakgroep Sociologie,
Grote Rozenstraat 31, 9712 TG Groningen. E-mail:
d.r.veenstra@ppsw.rug.nl

Tabel 1
Percentage scholen met verschillende scores op de vier succesmaten

	Verschil in variabelen		Verschil in methode	
	1A-1B	2A-2B	1A-2A	1B-2B
VBO				
geen	92	84	53	50
één cijfer	8	16	41	42
twee cijfers	0	0	5	5
meer dan twee cijfers	0	0	1	2
MAVO				
geen	100	42	45	34
één cijfers	0	42	45	45
twee cijfers	0	11	6	16
meer dan twee cijfers	0	5	5	5
HAVO				
geen	99	8	27	24
één cijfer	1	51	52	51
twee cijfer	0	32	21	22
meer dan twee cijfers	0	8	0	3
VWO				
geen	100	77	43	46
één cijfer	0	16	45	36
twee cijfers	0	7	10	15
meer dan twee cijfers	0	0	2	3

Legenda:

1A: methode Trouw - correctiefactoren Trouw

1B: methode Trouw - alternatieve correcties

2A: alternatieve methode - correctiefactoren Trouw

2B: alternatieve methode - alternatieve correcties

Tabel 2
Vershil en overeenkomst in scores tussen de succesmaten uit stap 1A en stap 2B,
per schooltype uitgedrukt in percentages

VBO						MAVO						
cijfer stap 2B						cijfer stap 2B						
	4	5	6	7	8		4	5	6	7	8	
cijfer stap 1A	4	<u>3</u>	3	1	1	0	4	<u>1</u>	2	3	1	0
	5	3	<u>7</u>	9	4	0	5	1	<u>5</u>	10	5	1
	6	2	7	<u>21</u>	5	1	6	3	8	<u>16</u>	9	2
	7	0	1	10	<u>14</u>	1	7	1	4	12	<u>11</u>	3
	8	0	0	1	4	<u>1</u>	8	0	1	1	1	<u>0</u>
HAVO						VWO						
cijfer stap 2B						cijfer stap 2B						
	4	5	6	7	8		4	5	6	7	8	
cijfer stap 1A	4	<u>2</u>	2	2	0	0	4	<u>0</u>	0	7	2	0
	5	3	<u>7</u>	10	8	0	5	2	<u>10</u>	12	5	0
	6	0	7	<u>16</u>	8	3	6	2	2	<u>21</u>	2	0
	7	3	8	8	<u>8</u>	2	7	0	2	12	<u>12</u>	5
	8	0	2	0	3	0	8	0	0	2	0	<u>0</u>

Legenda:

1A: methode Trouw - correctiefactoren Trouw

2B: alternatieve methode - alternatieve correcties

Tabel 3**Percentage scholen met een voldoende of onvoldoende in stap 1A en stap 2B**

stap 1A	stap 2B	VBO	MAVO	HAVO	VWO
onvoldoende	→ voldoende	15	20	20	26
voldoende	→ onvoldoende	10	17	20	6
voldoende	→ voldoende	58	55	48	54
onvoldoende	→ onvoldoende	16	9	14	12