

Dronkers' rapportcijfer voor de middelbare scholen: nauwkeurigheid, geldigheid en betekenis

C.M. van Putten en L.J.T. van der Kamp,
Rijksuniversiteit Leiden

Het doel van dit commentaar is met enige afstand en vanuit verschillende perspectieven de waarde vast te stellen van de rapportcijfers die Dronkers heeft toegekend in het dagblad Trouw (1997) aan de opleidingen van alle Nederlandse middelbare scholen. De centrale vragen bij het gebruik van rapportcijfers voor de prestaties van scholen zijn:

- Hoe komt een rapportcijfer tot stand en hoe nauwkeurig zijn de gegevens waarop rapportcijfers zijn gebaseerd?
- Wat betekent een rapportcijfer? Wat kun je op grond van een rapportcijfer voorspellen en hoe goed kun je dit doen?
- Met wat voor andere typen van meting en waardering zijn rapportcijfers te vergelijken?

Niet al deze vragen kunnen in het hierna volgende uitgebreid aan de orde komen. Wij zullen ons beperken tot het probleem van de nauwkeurigheid van de gebruikte gegevens en tot de wijze waarop het rapportcijfer tot stand kwam. Hierbij volgen wij de procedure van Dronkers op de voet. Bijzondere aandacht besteden wij aan het probleem van de betekenis van de rapportcijfers. Tot slot maken wij enkele opmerkingen over maatschappelijk nut en effect van het waarderingssysteem aan de hand van rapportcijfers.

1 De gebruikte gegevens

Een bekend methodologisch adagium is: 'Garbage in, garbage out'. Of om het wat positiever te formuleren: het resultaat van een berekening kan natuurlijk nooit beter zijn dan de kwaliteit van de gegevens die in de berekening gestopt worden.

1.1 Onwaarschijnlijkheden in de data

Uit de computer-output van Dronkers' berekeningen (1997) blijkt dat er onwaarschijnlijke gegevens in het databestand zitten. Wij noemen er enkele:

Het Trouw-overzicht laat twee VBO-opleidingen met een 1 als rapportcijfer zien. De output van Dronkers bevat steeds één VBO-opleiding met een cijfer van 1, maar er zijn ook ontbrekende gegevens of 'missing data'. In het derde leerjaar van deze opleiding zou 98% van de leerlingen tussentijds zijn uitgestroomd, en zou 54% zijn blijven zitten zonder de opleiding te verlaten. Dit kan natuurlijk alleen als deze percentages op twee verschillende opleidingen slaan. Maar dan nog zijn het percentages die onwaarschijnlijk zijn: of het zijn fouten in de data, of er is op de betreffende opleiding sprake van een zeer bijzondere situatie. Trouw (25 oktober 1997) vermeldt zelf bij een overzichtje van de 'slechtste scholen van 1996' dat de scholengemeenschap Harreveld (rapportcijfer 1) verbonden is aan een jeugdgevangenis.

Ook in de VWO-data komen onwaarschijnlijkheden voor: er is één VWO-opleiding met rapportcijfer 2, waarvan 61% van de leerlingen in het 4e leerjaar zou zijn blijven zitten, en 41% in de 5e tussentijds zou zijn uitgestroomd. Let wel: dit zijn alleen die gevallen in de output waarbij een gemiddelde was berekend over één opleiding en dus enkele gegevens van een afzonderlijke opleiding zichtbaar waren; er kunnen nog andere gevallen verscholen zitten in gemiddelden over meerdere opleidingen.

In geen geval mogen dit soort onwaarschijnlijke gegevens 'ongezien' in een berekening worden gestopt als je rapportcijfers aan individuele opleidingen wil toekennen. Uitval- en zittenblijfspercentages boven de 15 à 20% moeten op face-value gewantrouwd worden: het gaat hoogstwaarschijnlijk om bijzondere situaties (als opleidingen verbonden aan een gevangenis, opleidingen met veel asielzoekers,

* Met dank aan prof. P. van den Broek voor zijn opmerkingen en suggesties bij eerdere versies van dit commentaar.

of met een IVBO- of een VSO-afdeling), om fouten, of om getallen die er om administratieve of bureaucratische redenen staan. In één geval heeft Dronkers door correctie voor een groot aantal asielzoekers het rapportcijfer van de opleiding verhoogd van een 3 naar een 8 (Trouw, 22 november 1997).

1.2 Ontbrekende gegevens

De meest variabelen die Dronkers in zijn berekening stopt kennen een beperkt aantal 'missing data', enkele variabelen hebben echter zeer veel ontbrekende waarden (tot aan 61% toe). Er zijn in het algemeen tenminste twee redenen waarom gegevens kunnen ontbreken: zij zijn in principe wel voorhanden maar om een of andere reden niet geregistreerd, of zij zijn niet voorhanden en kunnen dus niet worden geregistreerd. Dit onderscheid is van belang voor de wijze waarop het probleem van missing data wordt onderzocht. Dronkers vervangt dergelijke 'missings' door het gemiddelde van de wel aanwezige gegevens. Weliswaar is dit een veelgebruikte procedure in de onderzoekspraktijk waar het gaat om groepsvergelijking. Maar de methodologische literatuur over het behandelen van ontbrekende gegevens (zie bv. Little & Schenker, 1995) leert dat we voorzichtig moeten zijn met de zogenaamde vervanging-door-het-gemiddelde procedure (zeker in gevallen waar meer dan 50% van de waarden ontbreken). Het is dan ook niet verantwoord om bij variabelen met 61% missings, waarden te genereren uit de 39% wel aanwezige gegevens: er is geen enkele garantie dat de aanwezige data representatief zijn voor de gehele groep. Het vervangen van zo veel 'missings' door het gemiddelde levert natuurlijk ook een heel extreme frequentieverdeling op: een enorme piek op het gemiddelde. Dergelijke variabelen kunnen dan ook beter buiten beschouwing worden gelaten.

De bovenstaande waarschuwing had betrekking op het onderzoek naar verschillen tussen groepen. Bij het rapportcijfer van Dronkers gaat het echter om het toekennen van scores aan afzonderlijke opleidingen. Dan is het toerekenen van gemiddelden aan een opleiding met een 'missing' niet zinvol: als de waarde die ontbreekt in feite bovengemiddeld is, dan wordt de opleiding benadeeld in het rapportcijfer, en omgekeerd treedt er bevoordeling op.

In de verdere berekening is het alleen bij het VWO zo dat variabelen met veel 'missings' een onbelangrijke rol spelen; bij de andere opleidingen worden ze gewoon meegewogen. Bij het VBO zijn de zittenblijf-variabelen voor het 2e leerjaar met 39% ontbrekende gegevens zelfs de belangrijkste constituenten van het rapportcijfer voor de VBO-opleidingen. Bij de HAVO-opleidingen heeft geen van de betrokken variabelen minder dan 9% 'missings'.

Dronkers is zonder correctie van onwaarschijnlijke gegevens en zonder het opvragen van ontbrekende gegevens met de inspectiedata aan de slag gegaan bij de berekening van een rapportcijfer per opleiding van elke school. Hoe zou iemand het vinden als een geneeskundige dienst een score openbaar maakt over zijn/haar gezondheidstoestand op basis van een aantal diagnostische gegevens, waarbij enkele ontbrekende uitslagen vervangen waren door gemiddelden (van de afdeling waar de betrokkene werkt) en waarbij enkele zeer merkwaardige uitslagen niet aan een nader onderzoek waren onderworpen?

1.3 Vergelijkbaarheid van de aanwezige gegevens

Maar ook op een deel van de gegevens die wel aanwezig zijn en die op het oog niet onwaarschijnlijk extreem zijn, valt het een en ander op te merken. Een belangrijke vraag is of die gegevens wel zinvol met elkaar te vergelijken zijn. Enkele voorbeelden:

Per opleiding zijn gemiddelde eindexamencijfers voor de vakken Nederlands, Wiskunde(-A) en Engels berekend. Omdat Nederlands voor alle eindexamenkandidaten verplicht is, zijn de opleidingen op dat gemiddelde te vergelijken, d.w.z. binnen VWO, HAVO, enz. Maar voor Engels en met name voor Wiskunde-A berust de vergelijkbaarheid op een wankel basis. Als opleidingen verschillen in de mate waarin leerlingen Engels of Wiskunde-A (of Wiskunde-B) als eindexamenvak kiezen, bestaat de kans dat een gemiddelde berust op een selectie van leerlingen uit de gehele opleiding, waardoor verschillen tussen gemiddelden aan uiteenlopende bronnen kunnen worden toegeschreven.

Een soortgelijk probleem doet zich voor bij MAVO en VBO waar Dronkers voor alle drie

de genoemde vakken tot een opleidingsgemiddelde komt door de waarden voor niveau-C en niveau-D te middelen. Dat gaat goed als op iedere opleiding deze beide niveaus in gelijke verhouding door leerlingen gekozen worden. Is dat niet het geval, en de kwaliteit van een opleiding zou kunnen blijken uit relatief meer van het zwaardere D-niveau, dan zijn dergelijke gemiddelden opnieuw niet goed meer te vergelijken. Zo berekent Dronkers geen Wiskundevariabele voor de VBO-opleidingen omdat er opleidingen zijn die geen eindexamenkandidaten met een C- of D-niveau hebben, hetgeen al een indicatie genoemd kan worden voor verschillen tussen opleidingen in dit opzicht.

De genoemde voorbeelden hebben betrekking op het welbekende methodologische fenomeen van selection bias. Hoe met verschillende vormen van selectiviteit om te gaan en welke modellen er zijn om dit probleem aan te pakken, is in de methodologische literatuur te vinden (zie bv. Winship & Mare, 1992, en Berk, 1983 voor een inleiding tot deze problematiek).

1.4 Allochtone leerlingen

In zijn berekeningen (zie straks paragraaf 2.2) probeert Dronkers rekening te houden met het percentage allochtone kinderen. Dit percentage is echter alleen per school bekend, en niet voor alle afzonderlijke opleidingen die door een school worden verzorgd. Dronkers neemt nu eenzelfde percentage allochtonen voor alle opleidingen van een school, terwijl bekend is dat allochtone leerlingen overwegend VBO en MAVO volgen. Het ongewenste effect hiervan is het sterkst voor VWO en HAVO: voor elke 12% allochtone leerlingen die een dergelijke opleiding 'extra' heeft omdat er zoveel in de VBO- of MAVO-tak van de school zitten, krijgt een HAVO- of VWO-opleiding een rapportcijfer dat een vol punt hoger is. Dat kost diezelfde school dan wel een driekwart punt voor zijn MAVO- en een half punt voor zijn VBO-opleiding, uitgaande van een school met opleidingen van ongeveer dezelfde omvang. Alleen voor categoriale scholen geldt dit bezwaar niet.

2 De bewerking van de gegevens

Dronkers distilleert een rapportcijfer per opleiding van iedere school uit een groot aantal gegevens van de inspectie over het eindexamen 1996 en het schooljaar 1995/96 (Dronkers in Trouw van 8 november 1997, pag. 12). De berekende rapportcijfers zijn op de volgende variabelen gebaseerd:

- het percentage geslaagden voor het eindexamen en de gemiddelde eindexamencijfers voor de vakken Nederlands, Wiskunde(-A) en Engels (4 variabelen, behalve voor het VBO waar zoals gezegd geen cijfer voor Wiskunde berekend kon worden);
- de percentages tussentijdse uitvallers, zittenblijvers die de opleiding blijven volgen, en zittenblijvers die de opleiding verlaten, voor alle leerjaren behalve het brugjaar en het eindexamenjaar.

Dit levert uiteenlopende aantallen variabelen op voor de verschillende opleidingstypen: 15 voor VWO (waarbij tussentijdse uitval in leerjaar 2 als gegeven ontbreekt), 13 voor HAVO, 10 voor MAVO, en 9 voor VBO (waarbij het cijfer Wiskunde ontbreekt). Laten wij dieper ingaan op de gevolgde werkwijze.

2.1 Het combineren van afzonderlijke variabelen

Als eerste stap combineert Dronkers de genoemde variabelen per opleidingstype tot een voorlopig rapportcijfer. De combinatie-techniek is Principale Componenten Analyse (PCA), een bekende beschrijvende en data-reducerende techniek: een groot aantal (met elkaar samenhangende) variabelen (met een grote hoeveelheid variantie) wordt vervangen door een klein aantal onafhankelijke componenten die zoveel mogelijk van die variantie moeten beschrijven. De eerste component beschrijft de meeste variantie, de tweede het meest van de resterende variantie, enz.. Dronkers kiest als voorlopig rapportcijfer de scores van de opleidingen op de eerste principale component; die beschrijft 'de feitelijke verschillen tussen scholen zo goed mogelijk' (Trouw, 8 november 1997). Hoewel dit een heel redelijke aanpak lijkt, kleven er toch enkele bezwaren aan en levert het in zijn uitvoering ook een aantal vervelende problemen op.

Het voornaamste bezwaar is dat een PCA niet de vraag beantwoordt wat de betekenis van de verkregen combinatiescore eigenlijk is. PCA geeft een spaarzame beschrijving van wat je erin stopt in de vorm van een reeks grootste gemene delers. Afgezien van het probleem dat er uitgegaan wordt van uiteenlopende aantallen variabelen per opleidingstype, is het de vraag hoe je die componenten moet interpreteren en hoeveel je er moet gebruiken. Een bijkomend – technisch – probleem is dat een PCA-oplossing mathematisch niet uniek is. Dit houdt in dat in het algemeen de gevonden componenten in een PCA-oplossing ter wille van de interpreteerbaarheid geroteerd mogen worden. Dit kan gebeuren zonder verlies van informatie in vergelijking met de oorspronkelijke oplossing. Dus welke PCA-oplossing te kiezen, geroteerd of ongeroteerd?

Dronkers interpreteert zelf het rapportcijfer als een cijfer voor het 'totale rendement van een opleiding van een school' en voor het al dan niet voldoende 'functioneren' van een school (in Trouw, 8 november 1997). Bij de tweede interpretatie verschuift de betekenis van het rapportcijfer van een samenvatting van rendementsvariabelen naar een indicatie voor de kwaliteit van een opleidingsorganisatie die aan bepaalde 'outputvariabelen' ten grondslag zou kunnen liggen.

De variabelen die Dronkers gebruikt zijn zonder twijfel belangrijke en relevante aspecten van het rendement van een school, maar zijn zelfs als men zich tot zoiets als kwantitatief rendement van een opleiding wil beperken, toch een te magere subset van na te streven onderwijs- en vormingsdoelen. Het is opmerkelijk dat geen enkele poging wordt gedaan tot een betekenisdefinitie of -analyse van 'functioneren' van een opleiding. Ook als men zich beperkt tot kwaliteit in de zin van 'hoe snel en hoe goed komen de leerlingen aan een diploma' (Agerbeek in Trouw, 8 november 1997) is een analyse vereist welke aspecten dit omvat en hoe men die moet combineren tot een rapportcijfer.

Het is ook opmerkelijk dat Dronkers niets doet met de uitkomsten die de PCA's nog meer opleveren dan de eerste component. Die eerste component beschrijft steeds maar een klein deel van de informatie die in de 9 à 15 variabe-

len opgeslagen ligt: dit varieert van slechts 17% (voor VWO), tot 20% (HAVO), 25% (MAVO) en 24% (VBO). Dat roept de vraag op welke informatie in de 2e resp. 3e componenten is samengevat en of een tweede rapportcijfer niet voor de hand had gelegen? Dronkers reduceert zijn data wel erg drastisch en de vraag is dan ook of deze vergaande data-reductie verantwoord is. Voor het VWO ligt de zaak nog weer iets anders omdat Dronkers hier niet de 1e maar de 2e component gekozen heeft (beschrijft maar 15% van de informatie). Deze keuze maakt duidelijk dat Dronkers op zoek was naar combinaties die een contrast opleveren tussen resultaat enerzijds (slaagpercentage en cijfers), en uitval en zittenblijven anderzijds. Bij het VWO kwam die combinatie pas op de 2e component naar voren met vooral slaagpercentage en Wiskunde-A tegenover late uitval (leerjaar 5 en ook 4). Op de 1e VWO-component, die dus de grootste verschillen tussen de VWO-opleidingen beschrijft, gaat met name het cijfer voor Engels samen met vroege uitval en zittenblijven (leerjaar 2 en 3). Wat betekent dit?

Vroege en late uitval zijn onderwijskundig duidelijk verschillende fenomenen die te maken kunnen hebben met het selectiebeleid van een opleiding. Uit de PCA-oplossing van de VWO-opleidingen blijkt dat vroege uitval niet systematisch samenhangt met late uitval. Dronkers verwaarloost dit onderscheid. Daardoor kunnen opleidingen met een hoog rapportcijfer toch een hoge mate van uitval en zittenblijven in leerjaar 2 en 3 hebben. Had dat onderscheid niet in het 'functioneren' betrokken moeten worden? Is het niet relevant voor ouders en leerlingen om bij de schoolkeuze te weten dat scholen zowel verschillen in de mate van de vroege als de late selectie van leerlingen?

Door een PCA per opleidingstype uit te voeren introduceert Dronkers een lastig probleem: PCA-oplossingen die niet zonder meer vergelijkbaar zijn, waardoor rapportcijfers ontstaan die per opleidingstype net weer iets anders betekenen:

– VWO: vooral slaagpercentage en cijfer Wiskunde-A, en ook Nederlands en Engels, tegenover zittenblijven (met en zonder vertrek) in leerjaar 5 en 4;

- HAVO: vooral slaagpercentage en cijfer Wiskunde-A, ook Nederlands (maar niet Engels), tegenover vooral zittenblijven met vertrek in leerjaar 3 en 4, en ook zittenblijven zonder vertrek in 3, 2, en 4;
- MAVO: vooral slaagpercentage en cijfers Nederlands en Wiskunde, ook Engels, tegenover zittenblijven met vertrek in leerjaar 3 en 2 en tussentijds vertrek in leerjaar 3;
- VBO: slaagpercentage en cijfer Nederlands (maar niet Engels), tegenover vooral zittenblijven (met en zonder vertrek) in leerjaar 2 en 3.

In ieder geval is duidelijk dat het rapportcijfer voor VBO-opleidingen een heel andere betekenis heeft dan voor de overige opleidingstypen: een hoog rapportcijfer in het VBO betekent vooral weinig zittenblijven en vertrek op het einde van het schooljaar. In de andere opleidingstypen slaat het rapportcijfer vooral op verschillen in slaagpercentages en gemiddelde cijfers. Het rapportcijfer voor HAVO's zegt weer niets over eindexamencijfers voor het vak Engels, in tegenstelling tot dat voor VWO en MAVO.

Samenvattend: zelfs met de beperkte gegevens die voorhanden waren, was het mogelijk geweest een rapportcijfer (of meerdere) te bepalen dat een meer doordachte en uniformere betekenis heeft, dan het cijfer van Dronkers. Wreekt zich hier de benadering vanuit een data-driven en puur kwantitatief-beschrijvende sociologie? Hoe het ook zij, ook data-analytisch blijven er vragen.

2.2 Correctie voor aanvangsverschillen tussen leerlinggroepen

Als tweede stap corrigeert Dronkers de PCA-scores voor hun samenhang met het percentage allochtonen, opnieuw per opleidingstype. Het achterliggende idee is als volgt: als twee opleidingen eenzelfde kwaliteit van 'functioneren' hebben, dan zal hun 'rendement' toch verschillen als zij met leerlingen van een ander gemiddeld niveau moeten werken (zie bv. Judd & Kenny, 1981). In een evaluatie-onderzoek dat opleidingen met elkaar vergelijkt op een aantal uitkomsten, streef je naar vergelijkbare groepen leerlingen die aan de opleidingen deelnemen. Omdat dat in de praktijk meestal niet

lukt, corrigeer je de uitkomsten voor dergelijke aanvangsverschillen tussen leerlinggroepen. Alleen als zo'n correctie slaagt, is het zinvol opleidingen met elkaar te vergelijken op hun uitkomsten. Cruciaal is de keuze van de correctiefactoren. Dronkers gebruikt het percentage allochtone leerlingen op een school als correctievariabele, niet omdat het de beste correctiefactor is. Het is de enige variabele die beschikbaar was; beter iets dan niets. Dan rest slechts de vraag hoe veel dat 'iets' is. Dronkers is daar optimistisch over: schoolprestaties einde Basisonderwijs, sociaal-milieu en allochtoon-zijn hangen sterk met elkaar samen, en worden goed gerepresenteerd door het percentage allochtone leerlingen op de opleiding. Daarmee zijn echter niet alle problemen opgelost.

De allochtone leerlingen zijn merkwaardig verdeeld over de Nederlandse opleidingen. Er is een beperkt aantal scholen met zeer grote aantallen allochtone leerlingen, tegenover grote aantallen scholen met betrekkelijk weinig of vrijwel geen allochtone leerlingen. De overwegend allochtone scholen zijn vooral te vinden in de grootste steden in het westen van het land. Waar weinig 'allochtoniciteit' is, kunnen zich echter aanzienlijke verschillen in aanvankelijk kennisniveau en in de sociale achtergrond van de leerlingen voordoen. Daarvoor wordt nu dus niet gecorrigeerd. Dat zou er een oorzaak van kunnen zijn dat de rapportcijfers van de opleidingen in de drie noordelijke provincies gemiddeld zo laag uitvallen.

Technisch gezien verloopt de correctie door van de verschillen tussen opleidingen in PCA-score, die verschillen af te trekken die samenhangen met het percentage allochtonen. De correlatie tussen beide variabelen ligt afhankelijk van opleidingstype tussen de 0,40 (VBO) en 0,49 (MAVO): hoe meer allochtone leerlingen, hoe lager de rendementscore. Vereist is wel dat die relaties lineair zijn en dat is hier niet te verwachten met zo'n scheef verdeelde 'allochtoniciteits'-variabele. Afhankelijk van de consequenties kan over- dan wel ondercorrectie optreden. De vraag is of het gebruik van lineaire data-analytische technieken (waar- onder PCA en correlatie) wel het meest geëigend is.

Voorts blijft een aantal kwesties met betrekking tot allochtoniciteit en correctie voor ver-

schillen tussen leerlinggroepen onopgelost:

- Speelt allochtoniteit alleen vooraf een rol (instroomverschillen) of ook tijdens de rit?
- Is een uiteenlopende correctie per opleidingsniveau gewenst?
- Zijn er geen andere correcties denkbaar voor omgevingsvariabelen zoals provincie, grote stad, en denominatie?
- En hoe staat het met correcties voor structuurvariabelen zoals schooltype? De scholen verschillen in het scala van opleidingen die zij verzorgen: van categoriale scholen (voor IVBO, VBO, MAVO of gymnasium), tot smalle scholengemeenschappen (VBO/MAVO, HAVO/VWO, MAVO/HAVO/VWO), tot brede scholengemeenschappen (VBO/MAVO/HAVO/VWO); ook heeft maar een deel van de VBO-opleidingen een IVBO-afdeling.

2.3 Een rapportcijfer voor rendement

Als derde stap wordt de gecorrigeerde PCA-score in een rapportcijfer omgezet lopend van minimaal 1 tot maximaal 10. Dat gaat als volgt:

- het gemiddelde rapportcijfer wordt per opleidingstype op 6 gesteld;
- getallen worden afgerond op hele cijfers (Dronkers: om schijn-exactheid te vermijden. Dit kan ten onrechte suggereren dat het afgeronde cijfer wel nauwkeurig is).
- scores onder de 1 worden op 1 gesteld; cijfers boven de 10 op 10 (om hoeveel gevallen ging het hier?)
- opleidingen met minder dan 50 leerlingen krijgen geen cijfer.

Dat het gemiddelde op 6 stellen en de rest tussen de 1 en 10 plaatsen een verkeerde beslissing is geweest, kan niet beter geïllustreerd worden dan met het dagblad Trouw zelf: pal naast een stuk van Dronkers met als kop: 'Schoolcijfer zegt niet dat scholen met 5 onvoldoende functioneren', bestaat Trouw het om te debiteren: 'De beste en de slechtste scholen van het land. Trouw-onderzoek Schoolprestaties. Nu in een speciale brochure' (advertentie in Trouw, 8 november 1997). Een 5-puntschaal, of een gemiddelde van 10, 20, of 50 was natuurlijk beter geweest, omdat getallen ten opzichte van zo'n gemiddelde geen automatische associatie met (on)voldoende oproepen.

3 Conclusie

De vraag is welk doel gediend wordt met een onderzoek naar de prestaties van opleidingen zoals gepresenteerd door Dronkers. In het algemeen kunnen het onderwijsveld, de scholen zelf alsmede het onderwijsbeleid op verschillende niveaus, en de ouders/leerlingen ermee gebaat zijn. Voor het onderwijsveld geven de resultaten van een dergelijk onderzoek de stand van zaken weer van, hopelijk, de kwaliteit van afzonderlijke opleidingen en scholen. Voor het onderwijsbeleid biedt het wellicht een handvat voor gerichte maatregelen. Voor ouders en leerlingen kunnen de resultaten behulpzaam zijn bij de vragen: Naar welke school zal ik mijn kind sturen? Tussen welke twee of drie scholen zal ik kiezen? De maatschappelijke relevantie van dergelijk onderzoek staat dan ook buiten kijf.

Een noodzakelijke voorwaarde is echter dat alle gevolgde methodologische en methodisch-technische beslistpunten en procedures inzichtelijk zijn en wetenschappelijk verantwoord worden. Uit het bovenstaande commentaar is duidelijk dat dit naar onze mening niet het geval is. Twijfels zijn gerezen bij de nauwkeurigheid van de gebruikte gegevens, bij de beslissing bepaalde variabelen op te nemen in de berekeningen (bv. allochtoniteit), terwijl er meerdere punten van kritiek zijn met betrekking tot de data-analyse. Bijzonder teleurstellend is dat Dronkers zo weinig aandacht heeft besteed aan de betekenis van het door hem berekende rapportcijfer. De betekenis van het nu geproduceerde rapportcijfer is onhelder, varieert per opleidingstype en heeft betrekking op een te klein deel van de beschikbare informatie.

Deze bezwaren tellen des te zwaarder omdat 'maatschappelijke dienstverlening', zoals Dronkers deze vorm van toegepast wetenschappelijk onderzoek pleegt te noemen, ditmaal betrekking had op afzonderlijke opleidingen en scholen, en niet op anonieme categorieën van scholen (zoals 'scholen in grote steden'). Het uitdelen van rapportcijfers op het individuele niveau van scholen vereist getallen die nauwkeurig zijn en waarvan de betekenis helder is.

Literatuur

- Agerbeek, M. (1997). Trouw-onderzoek schoolprestaties: antwoord op de emoties. *Trouw*, 8 november 1997, p. 12.
- Berk, R.A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review*, 48, 386-398.
- Dronkers, J. (1997). Schoolcijfer zegt niet dat scholen met 5 onvoldoende functioneren. *Trouw*, 8 november 1997, p. 12.
- Dronkers, J. (1997). SPSS-output. Amsterdam: SCO-Kohnstamm Instituut.
- Judd, C.M., & Kenny, D.A. (1981). *Estimating the effects of social interventions*. Cambridge, MA: Cambridge University Press.
- Little, R-J.A., & Schenker, N. (1995). Missing data analysis. In G. Arminger, C.C. Clogg & M.E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 39-75). New York/London: Plenum Press.
- Trouw (1997). Trouw-onderzoek schoolprestaties. De beste en slechtste scholen van Nederland. *Trouw*, 25 oktober 1997.
- Winship, C., & Mare, R.D. (1992). Models for sample selection bias. *Annual Review of Sociology*, 18, 327-350.

Auteurs

C.M. van Putten is universitair docent op het terrein van de multivariate data-analyse en het onderwijskundig onderzoek. Hij is verbonden aan de sectie Methoden & Technieken van de vakgroep Psychologie van de Rijksuniversiteit Leiden.

L.J.T. van der Kamp is hoogleraar methoden en technieken van psychologisch onderzoek, in het bijzonder op het terrein van de psychometrische testleer. Hij is verbonden aan de sectie Methoden & Technieken van de vakgroep Psychologie van de Rijksuniversiteit Leiden.

Correspondentie-adres: C.M. van Putten, Vakgroep Psychologie, Sectie Methoden & Technieken, Rijksuniversiteit Leiden, Postbus 9555, 2300 RB Leiden.
E-mail: Putten@rulfsw.leidenuniv.nl