

- Harvard University Press.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 1008, 480-498.
- Loveless, T. (2009). The use and misuse of international assessments. *The Brown Center Report on American Education*, 2(3), 8-18.
- Manen, H. van. (2009). *Goochelen met getallen*. Amsterdam: Boom.
- Pongratz, L. (2006) Voluntary self-control: education reform as a governmental strategy. *Education Philosophy and Theory*, 38, 471-482.
- Rindermann, H., & Ceci, S. J. (2009). Educational policy and country outcomes in international cognitive competence. *Perspectives on Psychological Science*, 4, 551-568.
- Schleicher, A. (2007). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess? *Journal of Educational Change*, 8, 349-357.
- Taykayama, K. (2008). The politics of international league tables: PISA in Japan's achievement crisis debate. *Comparative Education*, 44, 387-407.
- Wuttke, J. (2007). Uncertainties and Bias in PISA. In S.T. Hopmann, G. Brinek & M. Retzl (Eds.), *PISA zufolge PISA – PISA According to PISA. Hält PISA, was es verspricht? Does PISA Keep What It Promises?* (pp. 241-264). Wenen: Lit-Verlag.

Manuscript aanvaard: 23 januari 2011

Auteur

Sjoerd Karsten is als bijzonder hoogleraar werkzaam aan de Universiteit van Amsterdam.

Correspondentieadres: Sjoerd Karsten, Afdeling Child Development and Education, Faculteit der Gedrags- en Maatschappijwetenschappen, Universiteit van Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam. E-mail: s.karsten@uva.nl.

De maatschappelijke en wetenschappelijke waarde van internationale data over onderwijsprestaties

J. Dronkers

1 Inleiding

In deze bijdrage ga ik speciaal in op de maatschappelijke en wetenschappelijke betekenis van internationale onderwijsprestaties. Daarbij heb ik mij vooral gebaseerd op mijn persoonlijke ervaringen met het verrichten van secundaire analyses met de hieronder te bespreken internationale data over onderwijsprestaties en op het begeleiden van een aantal internationaal vergelijkende studies.

2 Maatschappelijke waarde

Crossnationale gegevens, die wetenschappers maar ook beleidsmakers in staat stellen onderwijsuitkomsten uit verschillende landen met elkaar te vergelijken, komen niet in de eerste plaats beschikbaar door wetenschapsinterne oorzaken, maar door een aantal maatschappelijke, politieke en technologische ontwikkelingen.

2.1 Maatschappelijke ontwikkelingen

Die maatschappelijke ontwikkelingen, die leiden tot meer crossnationale data en vergelijkingen, kunnen aangeduid worden met het gegroeide (toegekende) belang van *human capital*. Na de afloop van de Tweede Wereldoorlog werd in de meeste democratische en kapitalistische landen het belang van een goed opgeleide beroepsbevolking en de noodzaak van de mobilisatie van alle beschikbare talenten erkend. Ook werd vanaf die tijd de rol van de overheid voor de schepping van de condities voor een goedopgeleide bevolking en voor de ontginning van alle beschikbare talenten voluit erkend. Een goed opgeleide beroepsbevolking en het gebruik van alle talenten waren nodig voor de wederopbouw na de oorlog, en voor de onderlinge concurrentieverhoudingen tussen kapitalistische samenlevingen. Vooral als hun onderlinge economische barrières verlaagd werden,

zoals in de Europese Gemeenschap gebeurde. Economieën zouden niet meer concurreren door protectie van interne markten met hoge tariefmuren, maar door de kwaliteit van hun producten en dus door de kwaliteit en inzet van hun productiemiddelen, waaronder de opgeleide beroepsbevolking. Deze maatschappelijke noodzaak van een goede opleiding voor allen, leidde dan ook al snel tot internationale vergelijkingen, zowel door intergouvernementele als academische organisaties. De OESO (Organisatie van Economische Samenwerking en Ontwikkeling, de denktank van de westerse geïndustrialiseerde landen, beter bekend onder het Engelse afkorting OECD, gevestigd in Parijs) hield zich daarom al vroeg bezig met het beoordelen van de kwaliteit van de opleiding van die beroepsbevolking, naast hun meer strikt economisch studie- en advieswerk. Zo maakt de OESO al sinds de jaren zeventig evaluaties van de onderwijssystemen van de aangesloten landen. De OESO besloot ook, na de erkenning dat het aantal en het niveau van de nationale diploma's onbruikbaar is voor de vergelijking van de kwaliteit van opleidingen tussen landen (Martens, Rusconi, & Leuze, 2007), eind van de jaren negentig tot de crossnationale meting van het *human capital* van de toekomstige beroepsbevolking. Dat leidde tot de invloedrijke PISA-studies (Programme for International Student Assessment) van de OESO. Deze studies meten "how far students near the end of compulsory education have acquired some of the knowledge and skills that are essential for full participation in society". Deze omschrijving laat duidelijk zien dat het bij PISA gaat om de hoeveelheid kennis en vaardigheden, en dat de plaats waar die kennis en vaardigheden verworven zijn (familie, *peer*groep, verenigingen, school, etc.) secundair is. Omdat PISA in haar enquêtes aan andere socialisatiekaders dan de school weinig aandacht besteedt (het meest nog aan het ouderlijk gezin, maar de *peer*groep van de leerlingen blijft geheel buiten beschouwing), dreigt snel het gevaar dat variaties in bruikbare kennis en vaardigheden aan het onderwijs worden toegeschreven, terwijl de oorzaak heel goed kan liggen in een ander socialisatiekader.

Dat perspectief is verschillend bij de con-

current: de IEA (International Association for the Evaluation of Educational Achievement), die in 1958 startte met een bijeenkomst van een groep geleerden, onderwijspsychologen, sociologen en psychometrici in het UNESCO Institute for Education in Hamburg, en dat sinds die tijd data verzameld over onderwijsresultaten in specifieke vakken (taal: PRILS (Progress in International Reading Literacy Study), wiskunde en natuurwetenschappen: TIMSS (Trends in International Mathematics and Science Study), en burgerschap: ICCS (International Civic and Citizenship Education Study)). Deze veel meer onderwijskundig samengestelde organisatie wil internationale standaarden geven die beleidsmakers helpen de relatieve zwakte en sterke kanten van hun onderwijsstelsels vast te stellen. Maar zelfs de stichters van de IEA meenden al dat de meeste landen vergelijkbare omschrijvingen hebben van de optimale onderwijsresultaten en dat ze vooral verschilden in de middelen om dat te bereiken. De leerprestaties die de IEA op de verschillende domeinen meet, zijn nauwer verbonden met de inhoud van de feitelijke curricula in de aangesloten landen en de IEA meet dan ook veel nauwkeuriger dan PISA of in de bedoelde en aangeboden curricula wel de gemeten kennis en vaardigheden worden geleerd. Wel moet hierbij opgemerkt worden dat correcties voor verschillen of overlap in feitelijke curricula tussen landen nauwelijks invloed hebben op de posities die landen innemen in de internationale rangorde. Kennelijk maakt het niet zo veel uit wat er precies getest wordt aan schoolse kennis.

2.2 Politieke ontwikkelingen

De politieke ontwikkelingen, die leidden tot meer crossnationale data en vergelijkingen, kunnen aangeduid worden met het trefwoord legitimiteit. In moderne samenlevingen ontlenen overheden een deel van hun legitimiteit aan de kwaliteit van de door hen verzorgde voorzieningen. Onderwijs hoort zeker daarbij (Meyer, Boli, Thomas, & Ramirez, 1997). Een goed functionerend onderwijs legitimeert zowel de bestaande politieke structuren als de maatschappelijke ongelijkheid. Een in de ogen van de burgers 'eerlijke' selectie en hoge kwaliteit in het onderwijs legi-

timeert de ongelijke verdeling van mensen over de ongelijke posities in de maatschappij (beroep, inkomen, gezondheid, partners). In de loop van de 20^{ste} eeuw kreeg ‘eerlijke’ selectie de betekenis dat alleen cognitieve prestaties en motivatie van de leerling doorslaggevend mogen zijn in het onderwijs, terwijl aangeboren eigenschappen zoals ouderlijk milieu, geslacht en etnische herkomst geen rol meer behoren te spelen (de zogenaamde gelijke startkansen in het onderwijs). Overheden maar ook politiek betrokkenen hebben daarom behoefte aan informatie over de mate van ‘eerlijke’ selectie en de kwaliteit in hun onderwijs. Dat helpt overheden een beter beleid uit te zetten (het overnemen van de *best practice* van de buurlanden), een perspectief te hebben voor de mate van ‘oneerlijkheid’ in hun onderwijs (“in andere landen is het nog veel erger”), maar dat helpt ook protesterende burgers en partijen (“wij lopen ver achter bij”). Bovendien speelden de mate van ‘eerlijke’ selectie en kwaliteit in het onderwijs ook een rol tijdens de koude oorlog, toen zowel communistische als kapitalistische landen claimden het meest meritocratische en beste onderwijsstelsel te hebben. Dit belang van de ‘legitimiteit’ van het nationale onderwijsstelsel verklaart ook de gevoeligheid van overheden voor uitkomsten van internationale vergelijkingen, zowel de rangordeningen (Nederland scoort goed dankzij ons beleid) als de effecten van sommige achtergrondvariabelen (Nederland kent een sterk effect van ouderlijk milieu op onderwijsprestaties, zie OECD, 2004). Hoe meer nationale overheden moeten opereren in een geglobaliseerde omgeving, zoals de Europese Unie, waardoor hun handelingsbereik kleiner is geworden (invoering van de euro; Schengenakkoord voor vrij verkeer van personen), des te belangrijker hun legitimatie door een ‘eerlijk’ en goed onderwijsstelsel wordt. In dit verband is het nuttig erop te wijzen dat het Verdrag van Maastricht, waardoor de euro mogelijk werd, tegelijkertijd het onderwijs tot exclusieve verantwoordelijkheid van de lidstaten verklaarde. De enige taak, die de Europese Commissie ten aanzien van het onderwijs heeft, is die van informatieverzameling en -uitwisseling. Deze legitimiteit maakte in Duitsland de publicatie van de

PISA-data zo explosief. Onderwijs is in Duitsland een zaak van de deelstaten (*Länder*), niet van de federale overheid (*Bund*). Als gevolg van de politieke verdeeldheid tussen deze deelstaten over het onderwijs (in noordelijke deelstaten bevorderde een SPD-overheid een soort middenschoolontwikkeling, terwijl in zuidelijke deelstaten een CDU-overheid dat tegenhield) bestaat er nog geen nationale dataverzameling over de in het onderwijs geleerde hoeveelheid geleerde kennis en vaardigheden.¹ Dat betekende dat onderwijsresultaten tussen deelstaten nauwelijks onderling vergeleken konden worden, en veel deelstaten deden en doen alles om dat zo te laten. De nationaal representatief verzamelde PISA-data maakten plotseling een publieke vergelijking van de verschillen in onderwijsresultaten van deelstaten wel mogelijk. De relatief lage rangorde van Duitsland te midden van de andere OESO-landen bleek vooral te wijten te zijn aan het gemiddeld lage kennis- en vaardighedeniveau in de noordelijke deelstaten, hoewel in die deelstaten de effecten van achtergrondvariabelen op de onderwijsprestaties kleiner bleek dan in de zuidelijke deelstaten. Dit verlies van de legitimiteit van het onderwijs in de noordelijke deelstaten door deze PISA-data leiden tot een golf van onderwijs hervormingen (waaronder de invoering van een centraal examen in Noordrijn-Westfalen en Hessen), allen bedoeld om het vertrouwen in de ‘eerlijke’ selectie en de kwaliteit in het onderwijs terug te winnen.

2.3 Technologische ontwikkelingen.

De technologische ontwikkelingen, die leidden tot meer crossnationale data en vergelijkingen, hebben vooral betrekking op de invoering van de computer. Zowel de hardware als de software, maar ook de groei in gedigitaliseerde administratieve data, maken crossnationale vergelijkingen mogelijk, waarvan vroegere geleerden alleen maar konden dromen. Ook maakten deze technologische ontwikkelingen het mogelijk dat data een veel ruimere verspreiding krijgen dan ooit voor mogelijk werd gehouden. Zowel IEA als PISA stellen bijna alle data vrij beschikbaar voor onderzoekers, omdat geïnteresseerde onderzoekers ze van hun internet-

pagina's kunnen downloaden en hun eigen analyses uitvoeren. Het betekende echter ook dat het aantal 'ongecontroleerde' secundaire analyses van deze data tegenwoordig groter is dan vroeger, en dat in deze analyses geheel andere onderzoeksvragen aan de orde komen dan de oorspronkelijke verzamelaars en opdrachtgevers hadden gedacht.

3 Wetenschappelijke waarde

Hierboven blijkt al dat de crossnationale data niet ontstaan zijn om een bepaalde diepzinnige theorie te toetsen met behulp van onderwijsdata. Het doel van bijna alle belangrijke datasets van IEA en PISA/OESO is bepaalde beleidstheorieën met betrekking tot onderwijsstelsels (en haar onderdelen) van een empirische onderbouwing te voorzien. De wetenschappelijkheid van die onderbouwing is beperkt omdat onderwijsstelsels (en haar onderdelen) niet via reële experimenten met voldoende gecontroleerde condities gevarieerd kunnen worden. Als men uitsluitend experimenteel dubbelblind getoetste hypothesen als wetenschappelijk bewijs toelaat, zijn deze crossnationale data onbruikbaar. Maar datzelfde geldt dan voor grote en belangrijke wetenschapsgebieden zoals de biologie met de evolutietheorie of de astronomie met al haar waarnemingen voorbij het zonnestelsel. In deze en andere wetenschapsgebieden (en dat geldt ook voor de onderwijswetenschappen) zal men niet veel verder kunnen komen dan zorgvuldige veldobservaties, gecombineerd met uitgekende methoden om ruis en vertekeningen die voortvloeien uit veldobservaties (zoals selectiviteit; omgekeerde causaliteit; samenhangende factoren en processen) te neutraliseren. De hierboven genoemde wetenschapsgebieden, die met dezelfde beperkingen worden geconfronteerd, laten ook zien dat dit probleem niets met de tegenstelling tussen harde en zachte wetenschap heeft te maken. Als men zorgvuldige veldobservaties, gecombineerd met uitgekende methoden om ruis en vertekeningen te neutraliseren, accepteert als middel in de wetenschap dan kunnen crossnationale data bruikbaar zijn voor wetenschappelijke doeleinden.

De crossnationale data van IEA en PISA hebben een andere belangrijke beperking: het zijn dwarsdoorsneden op een bepaalde leeftijd van de leerlingen (PISA) of een bepaalde niveau (IEA) in plaats van longitudinale metingen van steeds dezelfde leerlingen. Dit is een belangrijke beperking voor onderwijsonderzoekers, omdat onderwijs een onderdeel is van de ontwikkeling van menselijke embryo's tot volwassen mensen. Het procesmatige, longitudinale karakter van deze socialisering (en waarvan onderwijs slechts het geïnstitutionaliseerde deel vormt) maakt dat dwarsdoorsneden op een bepaald moment van die ontwikkeling altijd beperkt en tot op zekere hoogte zelfs vertekend en dus misleidend zullen zijn. Deze beperking van alle crossnationale data tot dwarsdoorsneden komt in verschillende wetenschapsgebieden voor, maar is echter voor het onderwijs ernstiger dan voor bijvoorbeeld stemgedrag, omdat onderwijs en socialisatie bij uitstek een proceskarakter heeft, in tegenstelling tot stemgedrag.

Deze beperking wordt versterkt doordat zowel de huidige PISA- als de IEA-vragenlijsten weinig retrospectieve vragen aan de leerlingen bevatten (vragen over feit uit het verleden van de respondent). Nu is de betrouwbaarheid op retrospectieve vragen altijd discutabel, maar dat is geen reden ze dan maar niet te stellen. Een technische oplossing zou zijn meer historische gegevens uit de school- en onderwijsadministratie over de leerling te koppelen aan de antwoorden van de leerlingen. Op langere termijn is een dergelijke koppeling gemakkelijker door te voeren als gevolg van technologische ontwikkelingen (in Nederland zou het onderwijsnummer hiervoor bruikbaar zijn). Wel blijft er dan nog het ingewikkelde probleem om deze historische gegevens uit de school- en onderwijsadministratie internationaal vergelijkbaar te maken. Bovendien zullen niet alle deelnemende landen de noodzakelijke gegevens uit school- en onderwijsadministratie kunnen (ze bestaan niet) of willen (privacy) leveren.

Er zijn pogingen om deze beperking op te heffen door gebruik te maken van bepaalde technieken, meestal afkomstig uit de econometrie. Zo bestaat er de techniek van (*fuzzy*)

regression-discontinuity, die gebruik maakt van bepaalde breekpunten in de tijd (zoals de datumgrens bij de vorming van jaargroepen: 1 oktober in Nederland). In Nederland is deze techniek succesvol toegepast door Luyten (2006). Een andere techniek is de Differences-in-Differences-methode, waarbij twee dwarsdoorsneden van vergelijkbare leerlingpopulaties aan elkaar worden gekoppeld, maar waarbij de eerste, jongere dwarsdoorsnede nog niet is blootgesteld aan een bepaalde *treatment*, terwijl de tweede, oudere dwarsdoorsnede daar al wel is aan blootgesteld. Hanushek en Wössmann (2006) hebben deze techniek toegepast om de effecten van de mate van externe differentiatie in het voortgezet onderwijs vast te stellen. Een derde techniek is *propensity score matching*, waarbij geprobeerd wordt de ongemeten effecten van zelfselectie te elimineren, bijvoorbeeld schoolkeuze (Dronkers & Avram, 2010). Hoewel deze econometrische technieken tot op zekere hoogte de beperkingen van dwarsdoorsneden kunnen opheffen, vereisen ze vaak sterke, maar niet erg realistische, aannames waardoor de resultaten niet erg robuust zijn (Jakubowski, 2010) of zijn ze alleen bruikbaar voor heel specifieke, maar beperkte, onderzoeksvragen.

Hoewel het in principe mogelijk is dat IEA en PISA in de toekomst longitudinale gegevens in plaats van dwarsdoorsneden gaan verzamelen, is de kans daarop niet erg groot. Privacybelemmeringen en de enorme kosten van een dergelijke longitudinale meting maken het niet waarschijnlijk dat de financierende en subsidiërende overheden daarin geld willen steken. Bovendien hebben longitudinale data vanuit het perspectief van beleidsmakers een belangrijke beperking: ze hebben meestal betrekking op het onderwijsbeleid van een aantal jaren geleden en dus zijn de resultaten met deze 'verouderde' data niet relevant voor de huidige beleidskeuzen. Deze opvatting mag wellicht gebaseerd zijn op een – vanuit wetenschappelijk oogpunt – overschatting van het belang van het onderwijsbeleid voor de uitkomsten van onderwijsprocessen, maar dat verandert weinig aan de politieke realiteit. Daarom blijken de huidige dwarsdoorsneden van PISA en IEA de best mogelijke data om de effecten van

(delen van) onderwijsstelsels op geïnstitutionaliseerde socialisatie te meten.

Samenvattend ligt de kracht van deze crossnationale databestanden vooral bij de analyse van de effecten van onderwijsstelsels op het functioneren van het onderwijs bij het verwerven van kennis en vaardigheden. Ook zijn deze gegevens wel bruikbaar bij het meten van schooleffecten (de internationale data kennen een grotere variatie in schoolkenmerken dan nationale data), maar de meting van die schoolkenmerken is zeker niet optimaal. Bij de PISA-metingen zijn alleen de meningen van de directeur over het functioneren van de betrokken school en de leerkrachten beschikbaar, en zijn er geen middelen om de betrouwbaarheid van die mening van de directeur vast te stellen. Bij de IEA-metingen ligt het probleem precies anders om. De meningen van een aantal docenten over het functioneren van de betrokken school zijn bekend, maar het is onduidelijk of die gemiddeld het juiste beeld geven. Voor een wetenschappelijke analyse van onderwijsleerprocessen zijn de IEA- en PISA-data door hun dwarsdoorsnede karakter ongeschikt.

Zoals gebruikelijk zijn er beperkingen in de aard en het aantal verzamelde gegevens, maar dat is niet specifiek voor deze crossnationale dwarsdoorsneden. Wel is het mij opgevallen dat de IEA-datasets minder rijk zijn in het verzamelen van achtergrondkenmerken van de leerlingen dan de PISA-data. Wellicht ligt dit aan de onderwijskundige achtergrond van de belangrijkste deelnemers in IEA, maar ook de grotere politieke correctheid bij IEA kan een verklaring zijn. In de PISA- en OESO-organisatie spelen economen een grotere rol en die beroepsgroep heeft in vergelijking met onderwijskundigen minder last van politieke correctheid en een grotere interesse in de maatschappelijke omstandigheden van het onderwijs.

4 Problemen met de data

Door de hoge kosten van crossnationale data en door de noodzakelijke medewerking van veel overheidsinstanties is er sprake van een duidelijke politieke beïnvloeding van de data-

verzameling en -beschikbaarstelling, maar ook op de daarmee uitgevoerde analyses. Daarmee wil ik echter niet zeggen dat er bij nationale dataverzamelingen geen sprake zou zijn van politieke beïnvloeding. Hieronder zal ik een aantal voorbeelden behandelen, waarbij de nadruk op de PISA-data zal liggen omdat die data de grootste maatschappelijke impact gehad hebben.

Er is sprake van politieke invloed op data-verzameling en beschikbaarstelling. Het sterkste voorbeeld is Frankrijk: sinds de meting in 2003 zijn via PISA de antwoorden van de schoolleiders over de kenmerken van Franse scholen niet meer beschikbaar voor analyse door onafhankelijke wetenschappers. Australië en Canada geven het onderscheid tussen openbare en bijzondere scholen (juister geformuleerd tussen *public*, *private government-dependent* en *private government-independent* scholen) in hun landen niet in de publieke data set. Canada, Engeland², Frankrijk, Nederland³, USA, Zweden, en andere landen met veel immigranten vragen niet het specifieke geboorteland van leerlingen en ouders. Vanaf de meting in 2003 hadden de aan PISA deelnemende landen afgesproken dat elke land naar de belangrijkste geboortelanden van leerlingen en ouders zou vragen. Maar een groot aantal landen heeft dat niet gedaan: zij maken nog steeds alleen maar het onderscheid tussen wel of niet in het testland geboren, ondanks het feit dat het geboorteland van leerlingen of hun ouders meer variantie in onderwijsprestaties bindt dan hun land van bestemming (Heus & Dronkers, 2010; Levels, Dronkers, & Kraaykamp, 2006). In de meting van 2006 is de gezinsvorm, waarin de leerling op dat moment opgroeit (gezin met beide ouders, moedergezin, vadergezin, etc.) verdwenen als een van de achtergrondkenmerken, ondanks het feit dat gezinsvorm veel meer variantie verklaart dan de meeste schoolkenmerken (bijvoorbeeld het aantal computers in de school). Maar in de meting van 2009 is het weer terug.

Wie echter denkt dat de data van IEA minder beperkingen kennen, omdat dit een organisatie van wetenschappers is, komt bedrogen uit. Aan de meting van burgerschap (1999) doet Nederland niet mee. Toen vonden het Nederlandse Ministerie van Onder-

wijs en ook de vele belangengroepen die altijd de mond vol hebben over de maatschappelijke taak van het onderwijs het niet nodig om daar meer over te weten⁴. De opstellers hebben in de IEA-studies alle gezinsvormen op een veel te simpele driedeling geperst (twee ouders, inclusief stiefouder), moedergezin of vadergezin. Naar het specifieke geboorteland van leerlingen en hun ouders wordt in het geheel niet gevraagd. Ook de godsdienst van de leerlingen of de ideologische, religieuze of pedagogische achtergrond van de scholen ontbreekt in het IEA-onderzoek naar burgerschap. Zelfcensuur is vaak effectiever dan het afdwingen van politiek correcte opvattingen.

Het zal niet verbazen dat een dergelijke politieke beïnvloeding tot politiek correcte rapporten leidt. Het beste voorbeeld is het OESO-rapport "Where immigrants succeed. Pisa 2003" (OECD, 2006). In dit officiële rapport worden de onderwijsresultaten van leerlingen met een immigrantenachtergrond in de deelnemende PISA-landen geanalyseerd met behulp van de meting in 2003 van PISA. Dat was de eerste PISA-meting waarin een 15-tal deelnemende landen naar het specifieke geboorteland van leerlingen en ouders hadden gevraagd. Maar daarvan is in het rapport zo goed als niets terug te vinden. Leerlingen met een immigrantenachtergrond in Finland worden zonder meer vergeleken met leerlingen met een immigrantenachtergrond in Duitsland, ondanks het feit dat eerstgenoemde hoofdzakelijk uit Zweden en Rusland (waaronder het taalverwante Estland) afkomstig zijn, terwijl de laatstgenoemde uit Turkije, Joegoslavië, etc. De conclusie van dit rapport, dat leerlingen met een immigrantenachtergrond het beter doen in het Finse onderwijs dan in het Duitse, is dus misleidend, omdat hier prestaties van leerlingen uit zeer verschillende herkomstlanden vergeleken worden, zonder dat de kenmerken van die landen er bij betrokken zijn. Ook kregen de schrijvers van dit rapport van hun opdrachtgevers een feitelijk verbod om een analyse uit te voeren van de relatie tussen het in de deelnemende landen gevoerde beleid ten aanzien van leerlingen met een immigrantenachtergrond en hun scores op de kennis en vaardigheden toetsen. Merkwaardigerwijs zijn

er uit de onderzoekswereld maar weinig protesten te horen geweest over deze gang van zaken met dit OESO rapport en is het nog steeds invloedrijk.

Maar de crossnationale data van IEA en PISA hebben twee belangrijke kenmerken waardoor de politieke beïnvloeding door nationale overheden geen ernstige gevolgen hoeft te hebben. Het ontbreken van een land binnen een internationale vergelijking van bijvoorbeeld Europese onderwijsstelsels is niet leuk, maar ook niet ernstig. Als een bepaald effect in alle Europese landen gevonden wordt (bijvoorbeeld een positief effect van *private government-dependent* scholen op onderwijsprestaties; zie Dronkers & Robert, 2004) en dat effect is min of meer even groot in al die Europese landen, dan is het onwaarschijnlijk dat dit effect in Frankrijk niet zou bestaan. Met andere woorden, het crossnationaal karakter van deze data biedt voldoende mogelijkheden om nationale politieke beïnvloeding te omzeilen, want het is in de meeste gevallen onwaarschijnlijk dat het onderwijs in het ene land een heel ander effect heeft dan in buurlanden of in politiek-economisch verwante landen (zoals de OESO-landen)⁵. Een tweede kenmerk waardoor de gevolgen van politieke beïnvloeding beperkt kan blijven, is de onmiddellijke terbeschikkingstelling van de data. Dat vergroot de mogelijkheden om politiek correcte rapporten te controleren en eventueel te corrigeren, maakt het mogelijk dat politiek ongewenste vragen toch empirisch worden aangepakt door onafhankelijke onderzoekers, en biedt de mogelijkheid om geheel nieuwe vragen aan te pakken. Wat dat betreft voldoen de IEA- en de PISA-data goed aan een belangrijke eis binnen het wetenschappelijk bedrijf: de publieke beschikbaarheid van de data. De meeste schoolloopbaandata en andere onderwijsdata in Nederland voldoen in veel mindere mate aan dit criterium. Meestal is eerst toestemming van het CBS of een andere dataverzamelaar nodig, voordat een onafhankelijke wetenschapper aan de slag mag. Bovendien verloopt er vaak een lange tijd tussen de eerste publicatie en de beschikbaarstelling van data, teneinde de (onderzoeks-)belangen van dataverzamelaar maar ook opdrachtgever te beschermen.⁶

5 Onzinnige bezwaren

In de twee andere bijdragen over crossnationale data over leerprestaties komen een groot aantal problemen met deze data aan de orde. Uit het voorafgaande zal reeds duidelijk zijn dat die problemen reëel zijn. Maar soortgelijke problemen komen in veel nationale data ook voor, zonder dat er een haan naar kraait. Blijkbaar zijn de belangen bij de uitkomsten van crossnationale vergelijkingen groter, en wordt er aan die problemen plotseling zwaarder getild. Gelukkig geldt zowel voor IEA als PISA het spreekwoord “De honden blaffen, maar de karavaan trekt verder”, want een aantal bezwaren zijn onzinnig. Daarvan wil ik er drie hier bespreken.

1. De PISA-scores komen tot stand door bedrog. Scholen sturen hun slechtste leerlingen weg tijdens de PISA-testen en laten alleen hun beste leerlingen meedoen. Veel empirische bewijzen worden meestal niet aangedragen, maar laten wij voor het gemak aannemen dat scholen inderdaad alleen de betere leerlingen laten deelnemen. Als alle scholen dat in de zelfde mate doen is dat voor de internationale vergelijking van leerprestaties niet bezwaarlijk. In alle landen worden door dit bedrog de leerprestaties in dezelfde mate overschat. Het zou alleen bezwaarlijk zijn als in het ene land de mate van bedrog door scholen groter zou zijn dan in een ander land. Als dat waar zou zijn, is de logische conclusie dat Finse scholen het best zijn in dit bedrog. Ook zouden scholen in Noord-Italië beter zijn in dit bedrog dan scholen in Zuid-Italië. Het lijkt allemaal niet erg waarschijnlijk. Het is plausibeler om uit te gaan van de aanname dat scholen in de OESO-landen een vergelijkbaar niveau van bedrog hebben. Wel kunnen de verschillen in de omvang van het speciaal onderwijs in de deelnemende landen de vergelijkbaarheid beïnvloeden. Leerlingen in het speciaal onderwijs behoren niet tot de onderzoekspopulaties van PISA en IEA, maar de percentages leerlingen in het speciaal onderwijs variëren wel van land tot land. De hoge score van Nederland in de PISA-meting van 2000 lijkt te wijten aan het feit dat

leerlingen op lom- en mlk-scholen niet meedeeden (Knecht-Van Eekelen, Gille, & Van Rijn, 2007). Ook kan de feitelijke deelname van 15-jarigen aan het onderwijs niet in overeenstemming zijn met de leerplichtwetgeving in een land. Dit verschil tussen de juridische verplichting tot schoolbezoek en het werkelijke volgen van onderwijs is groot in deelnemende landen zoals Brazilië en Turkije, en in mindere mate ook voor andere landen (OECD, 2007).

2. De uitkomstverschillen tussen landen komen door de onderwijsprestaties van leerlingen met een immigratieachtergrond. Het niveau van het Duitse onderwijs is daardoor in PISA en IEA onderschat. Dat lijkt gezien het niet opnemen van het herkomstland (zie hierboven) een plausibel argument, en inderdaad wordt het verschil in de gemiddelde Duitse en Finse score kleiner, als men rekening houdt met de percentages en de verschillen in herkomst van migrantenleerlingen in die landen. Maar het percentage leerlingen met een immigrantenachtergrond is te klein⁷ en de verschillen in scores tussen de inboorlingen zijn te groot, om een substantiële verandering in de onderlinge rangorde te veroorzaken.
3. Uitkomsten voor taal zijn heel anders dan voor wiskunde of natuurwetenschappen. Het lijkt wellicht een voor de hand liggend bezwaar, maar dit wordt niet gesteund door de feitelijke uitkomsten met PISA-data. Er blijkt tussen de drie domeinen (taal, wiskunde, natuurwetenschappen) een grote gezamenlijke variantie te bestaan. Blijkbaar meten de testen voor deze drie domeinen in eerste instantie de manifeste intelligentie van leerlingen en pas in tweede instantie specifieke kennis en vaardigheden in een bepaald leerdomein. Deze gezamenlijke variantie is op het nationaal niveau nog sterker dan op het individuele niveau. Rindermann en Ceci (2009) hebben laten zien dat noch de inhoud van de toets (naar vakgebied), de aard van de getoetste kennis (curriculum of minder curriculum georiënteerd), het jaar van de meting of de onderzoekspopulatie (basisschool- of VO-leerlingen) veel

uitmaken voor de internationale rangorde⁸. De IEA-data laten een dergelijke conclusie niet toe, omdat IEA de testen voor de verschillende leerdomeinen niet bij dezelfde leerlingen afneemt. Sinds de meting in 2003 doet PISA dat wel en toen pas werd de grote gezamenlijke variantie zichtbaar. Omdat PISA burgerschapkennis en -vaardigheden niet meet, weten wij niet of deze overlap in variantie ook voor burgerschap geldt. Maar de effecten van de gebruikelijke achtergrondkenmerken zijn bij burgerschap niet erg verschillend van die van taal, wiskunde of natuurwetenschappen.

6 Gebruik van crossnationale data in Nederland

Alvast in Nederland is er onvoldoende wetenschappelijk toezicht op de PISA-dataverzameling. Alleen het Nederlandse Ministerie van Onderwijs is verantwoordelijk voor de dataverzameling (die in opdracht wordt uitgevoerd door het Cito) zonder overleg met onafhankelijke wetenschappers. Dat wreekt zich zowel bij de beslissingen over de invulling van de vragen (het beste voorbeeld hiervan is het niet-opnemen van het geboorteland in de metingen van 2003 en 2006⁹). Bij IEA lijkt de academische inbreng groter, maar ook hier heeft het ministerie van onderwijs een grote rol (lid van het *standing committee* en nationaal vertegenwoordiger in de algemene IEA-vergadering).

Daar komt bij dat verschillende politici (zoals Ritzen of Netelenbos) verwijzen naar de hoge scores van Nederlandse scholieren in de internationale rangordes van IEA en PISA om het succes van hun politiek te bewijzen. Maar zij hebben nooit de opdracht gegeven voor een onafhankelijk onderzoek naar de verklaring van de relatief hoge scores van Nederlandse scholieren in vergelijking met die van de andere Europese landen. Blijkbaar vonden deze politici het niet nodig hun gelijk empirisch te onderbouwen, of wilden zij niet weten waaraan Nederlandse leerlingen die hoge score te danken hebben. Daarom stelden zij geen enkele beurs voor postdocs of aio's beschikbaar voor een dergelijke secun-

daire analyse van de PISA- of IEA-data.

De grote invloed die nationale regeringen hebben op IEA en PISA is waarschijnlijk de onvermijdelijke prijs die de wetenschappelijke wereld moet betalen voor deze kostbare dataverzameling. Maar het rechtvaardigt niet de afwezigheid van wetenschappelijk betrokkenheid bij de dataverzameling en de analyses. Echter wij kunnen niet alleen de overheid de schuld hiervan geven. Ook de beroepsvereniging is op dit punt niet actief genoeg. Ik herinner mij geen serieuze aandacht aan internationale vergelijkingen op bijeenkomsten van de VOR. De VOR heeft nooit geprotesteerd tegen de machtspositie van het Ministerie van Onderwijs in PISA en IEA. Ook bestaat er geen afzonderlijk NWO- of PROO-programma meer voor de analyse van de crossnationale data van IEA of PISA.

Het is daarom niet verbazingwekkend dat de wetenschappelijke oogst van analyses van de IEA- en PISA data, die verder gaan dan een beschrijving van de primaire uitkomsten tegenvalt. Onderwijskundigen en sociologen, die – gegeven de mogelijkheden van crossnationale data voor hun belangrijkste onderzoeksvragen – het voortouw hadden moeten nemen, hebben het vaak laten afweten. Die rol lijkt overgenomen te zijn door economen, die de PISA- en de IEA-data ontdekt lijken te hebben. Zelfs in Nederland is dit goed te illustreren: het door economen beheerste Centraal Planbureau heeft meer met de PISA-data gedaan dan het door sociale wetenschappers gedomineerde Sociaal en Cultureel Planbureau. Er is dus voor Nederlandse onderzoekers nog een wereld te winnen.

Noten

- 1 Men is pas afgelopen jaar begonnen met een dergelijke nationale dataverzameling, die door zowel de *Bund* als de *Länder* betaald wordt, maar waarbij rechtstreekse vergelijking van onderwijsresultaten van deelstaten niet toegestaan is.
- 2 Schotland vraagt wel naar geboorteland, maar dat komt omdat het Schotse onderwijs buiten de eeuwigdurende Unie met Engeland is gebleven, net als de Schotse rechtspraak.
- 3 Vanaf de PISA-wave 2009 vraagt Nederland

wel naar het specifieke geboorteland van leerlingen en ouders, dankzij een intensieve lobby.

- 4 Nederland doet wel mee aan de ICCS-wave 2009.
- 5 Het is even waarschijnlijk dat in een bepaalde school het normale positieve effect van ouderlijk milieu op onderwijsprestaties niet zou bestaan maar dat op die bepaalde school juist een tegengesteld negatief effect zou voorkomen.
- 6 Overigens zijn de extra data en variabelen die Duitsland in elke PISA meting toevoegde, nog slechter bereikbaar voor onafhankelijke onderzoekers. Een gedetailleerd onderzoeksvoorstel wordt anoniem gerefereerd, en afwijzing is een realistische uitkomst, bij voorbeeld omdat analyse van effecten van herkomstlanden in Duitsland nog uit den boze is. Als een andere onderzoeker al met een verwante analyse bezig is, wordt er ook geen toestemming geven voor een andere secundaire analyse. Analyses die neerkomen op een vergelijking van onderwijsstelsels van deelstaten zijn ook niet toegestaan. Het kan dus altijd nog erger dan in Nederland. De webpagina van dit Duitse data-archief geeft meer (maar te optimistische) informatie over het verwerven van de Duitse PISA-data: [www.iqb.hu-berlin.de\FDZ](http://www.iqb.hu-berlin.de/FDZ).
- 7 PISA en IEA meten door het gebruik van geboorteland van leerlingen en ouders bovendien alleen maar eerste en tweede generatie, en de derde generatie blijft onzichtbaar. Op de lange termijn zal men toch moeten overgaan tot de meting van subjectieve etnische identiteit, zoals ook in de VS gebruikelijk is.
- 8 Als dit juist is, vereist dit wel een nieuwe door-denking van de validiteit van de taal, wiskunde en natuurwetenschappen toetsen. Wat meten wij met de PISA- en IEA-toetsen? Manifeste intelligentie? (Rindermann & Ceci, 2009).
- 9 Het geboorteland is in de meting van 2009 wel gemeten in Nederland, maar dat is het gevolg van langdurig lobbyen.

Literatuur

- Dronkers, J., & Avram, S. (2010). A cross-national analysis of the relations of school choice and effectiveness differences between private-

- dependent and public schools. *Educational Research and Evaluation*, 16, 151-175.
- Dronkers, J., & Robert, P. (2004). De effectiviteit van openbaar en bijzonder onderwijs: een crossnationale analyse. *Mens en Maatschappij*, 79, 170-192.
- Hanushek, E., & Wössmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal*, 11, C63-C76.
- Heus, M. de, & Dronkers, J. (2010). De schoolprestaties van immigrantenkinderen in 16 OECD-landen. De invloed van onderwijsstelsels en overige samenlevingskenmerken van zowel herkomst als bestemmingslanden. *Tijdschrift voor Sociologie*, 30, 260-294
- Jakubowski, M. (2010). Institutional tracking and achievement growth: Exploring Differences-in-Differences approach to PIRLS, TIMSS, and PISA data. In J. Dronkers (Ed.). *Quality and inequality of education. Cross-national perspectives* (pp. 41-82). New York: Springer.
- Knecht-van Eekelen, A., Gille, E., & Rijn, P. van. (2007). *Resultaten PISA-2006. Praktische kennis en vaardigheden van 15-jarigen. Nederlandse uitkomsten van het OESO Programme for International Student Assessment (PISA) op het gebied van natuurwetenschappen, leesvaardigheid en wiskunde in het Jaar 2006*. Arnhem, Nederland: CITO.
- Levels, M., Dronkers, J., & Kraaykamp, G. (2006). Het belang van herkomst en bestemming voor de schoolprestaties van immigranten. Een crossnationale vergelijking. In F. van Tubergen & I. Maas (red.). *Allochtonen in Nederland in internationaal perspectief* (pp. 137-160). Amsterdam: Amsterdam University Press.
- Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95. *Oxford Review of Education*, 32, 397-429.
- Martens, K., Rusconi, A., & Leuze, K. (Eds.) (2007). *New arenas of education governance – The impact of international organisations and markets on educational policymaking*. London: Palgrave
- Meyer, J. W., Boli, J., Thomas, G. M., & Ramirez, F. O. (1997). World society and the nation state. *American Journal of Sociology*, 103, 144-181.
- OECD. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.
- OECD. (2006). *Where immigrant students succeed. Pisa 2003*. Paris: OECD.
- OECD. (2007). *PISA 2006 Science Competencies for Tomorrow's World. Analysis*. Paris: OECD.
- Rindermann, H., & Ceci, S. J. (2009). Educational policy and country outcomes in international cognitive competence. *Perspectives on Psychological Science*, 4, 551-568.

Manuscript aanvaard: 23 januari 2011

Auteur

Jaap Dronkers is als hoogleraar werkzaam aan het Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA) van de Universiteit Maastricht.

Correspondentieadres: Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA), Universiteit Maastricht, Postbus 616, 6200 MD Maastricht.
E-mail: j.dronkers@maastrichtuniversity.nl