

Pisa en zijn verreregaande ambities

Roger Standaert¹

Nu het stof van de PISA-explosie 2015 wat is gaan liggen, is de tijd rijp om eens op wat afstand ernaar te kijken. Het is immers merkwaardig hoe de driejaarlijkse afkondiging van de PISA-resultaten wereldwijd in allerlei schoolsystemen tot een mediahoogfeest leidt. De OESO laat niets ongemoeid om via een sterk gestuurde campagne schoolsystemen, niet altijd zonder enige betutteling, te informeren over hun al dan niet verbeterde of verslechterde prestaties. De vereenvoudiging van een ongelooflijke hoeveelheid data in drie cijfers voor wetenschappen, respectievelijk wiskunde en leesvaardigheid maakt het ook mogelijk om voor een kritiekloos publiek een hitparade van onderwijssystemen te presenteren.

Tijd dus om wat vragen te stellen:

- Is het vergelijken van complexe onderwijssystemen wel mogelijk?
- Hoe worden de gegevens verzameld?
- Hoe betrouwbaar zijn de rankings die daarop zijn gebaseerd?
- Gaat het om de juiste vraagstellingen?
- Is het onderwijs wel verantwoordelijk voor de resultaten of zijn het eerder contextuele, historische en culturele karakteristieken rond een schoolsysteem?

Daarover gaat deze bijdrage. Om misverstanden te vermijden: er is niets mis met het leren van andere onderwijssystemen. Wel mag de vraag worden gesteld voor welk doel die vergelijkingen worden gemaakt en, uiteraard ook, of, wat in een ander systeem werkt, zomaar kan functioneren in je eigen onderwijssysteem?

I. Waar gaat het om?

Het Programme for International Student Assessment (PISA) is een project van de OESO, een groepering van 34 welvarende economieën in de wereld. Het is een organisatie die gericht is op economische vooruitgang. Ze is met die bedoeling opgericht in 1961.

De belangstelling van de OESO voor onderwijs dateert van het einde van de jaren 1990. Het economisch denken met indicatoren kreeg toen een zijweg in het onderwijs, oorspronkelijk vooral in de Verenigde Staten. Daar zat men sinds de Spoetnikcrisis van 1957, in volle Koude Oorlog, met een zware kater inzake achterstand op het Oostblok. De oorzaak werd gezocht in het lage niveau van het onderwijs. Zware investeringen waren daarom noodzakelijk. Maar die investeringen moesten wel duidelijk rendement opleveren, zodat de weg van tests moest worden opgegaan. Dat paste naadloos binnen de tradities van testpsychologie en het onderliggende behaviorisme in de Verenigde Staten. Vanaf 1969 werden steekproeven genomen van de prestaties op school voor de leeftijden van 9, 13 en 17 jaar. Na een reeks vernietigende rapporten over de bedenkelijke kwaliteit van het onderwijs in de VS werd in 1984 het verzamelen van de resultaatsgegevens voor alle deelstaten sterk gesystematiseerd. Vanaf 1990 werden de vergelijkingen tussen de deelstaten gepubliceerd.

In 1987 kwam de problematiek van het verzamelen van statistische gegevens op de agenda van de OESO. In 1992 verscheen de eerste uitgave van *Education at a Glance* waarin statistische gegevens rond een reeks indicatoren voor alle OESO-landen werden verzameld. Als grote speler (en als grote financier) binnen de OESO eisten de VS in 1998 dat ook de OESO zich zou gaan bezighouden met het meten van de onderwijskwaliteit binnen de OESO. De historicus TRÖHLER stelt dat de VS zelfs dreigden met het terugtrekken van hun lidmaatschap als de OESO daarin

¹ Ereprofessor comparatieve pedagogiek UGent.

niet zou meegaan. Er was weerstand bij de meeste van de 28 landen van de OESO die het nut en de haalbaarheid betwijfelden van dergelijke internationale vergelijkingen. Men vond, aldus TRÖHLER, dat er een groot gevaar was voor simplificaties en het fout voorstellen van de systemen binnen de lidstaten. Maar de impact van de VS was zo sterk dat de weg werd geplaveid voor het eerste PISA-rapport in 2001 (TRÖHLER, 2013).

Sindsdien wordt om de drie jaar een PISA-rapport afgeleverd over prestaties van 15-jarige leerlingen voor wetenschappelijke geletterdheid, wiskunde en leesvaardigheid. Meestal wordt ook bijkomend een thema toegevoegd. In 2003, 2012 en 2015 ging het zo om 'problem solving'. Het aantal deelnemende landen steeg constant met uitschieters van 65 onderwijssystemen in 2012 en 72 in 2015.

PISA pretendeert niet uit te gaan van de leerinhouden of programma's in de verschillende landen, maar van nuttige competenties die in deze tijd nodig zijn en die eventueel ook buiten de school kunnen worden verworven. Vele vragen uit de toetsen zijn dan ook gekaderd binnen een verbaal toegelichte situatie. Bij de toetsing worden ook een aantal leerlingkenmerken in rekening gebracht, vooral verzameld via vragenlijsten. Dat geldt ook voor schoolkenmerken, waarvoor men gebruikmaakt van vragenlijsten bij directies.

De resultaten worden statistisch verwerkt via het wegen van de vragen zodat per toets vaardigheidsniveaus kunnen worden afgebakend, vertrekkend van een gemiddelde en een standaarddeviatie. De resultaten worden dus ingedeeld in niveaus. In 'PISA 2015' worden zeven niveaus onderscheiden (voorheen zes). Bijvoorbeeld voor wetenschappen als volgt:

- niveau 6: 708 punten of hoger;
- niveau 5: 633-708;
- niveau 4: 559-633;
- niveau 3: 484-559;
- niveau 2: 410-484;
- niveau 1a: 335-410;
- niveau 1b: 261-355.

Alle data van PISA worden vrijgegeven voor onderzoeksdoeleinden, zodat wetenschappelijke instituten en andere belanghebbenden daarop verdere analyses kunnen uitvoeren. Daarvan wordt dan ook overvloedig gebruikgemaakt. Maar daarmee verwatert ook de controle op die analyses en bewerkingen. Het gevaar van 'cherry picking' is dan niet denkbeeldig.

II. Vergelijken is nuttig

Deze bijdrage stelt een aantal kritische vragen bij het 'hype'-karakter van de PISA-projecten. Dat betekent niet dat het

vergelijken van onderwijssystemen en hun karakteristieken niet zinvol zou zijn. Daarom is het nuttig hierover stelling in te nemen.

Zoals dat het geval is bij alle wetenschappen, is er ook bij het vergelijken van onderwijssystemen een evolutie te onderkennen. De belangrijkste bedoeling is altijd geweest om te leren van het buitenland en na te gaan hoe men in het buitenland onderwijsproblemen oplost. Daarbij is steeds de gedachte aanwezig dat een onderwijssysteem ingebed zit in een geschiedenis en in een culturele context, die het onmogelijk maakt structuren, aanpakwijzen en oplossingen zomaar over te nemen. Het is vergelijkbaar met het transplanteren van organen. Een orgaan dat getransplanteerd wordt, moet compatibel zijn met het organisme van de patiënt. Zo niet, wordt het afgestoten. Het is daarom lichtelijk naïef oplossingen vanuit het buitenland zomaar over te nemen. Er is altijd een aanpassing nodig om wat in het buitenland succes heeft, ook te doen slagen in het eigen onderwijssysteem.

Zo is het gekende 'duaal systeem' in Duitsland, waarbij vaklui opgeleid worden in een leerlingstelsel in nauwe samenwerking met het bedrijfsleven, niet zomaar over te plaatsen naar ons land. Het leerlingstelsel (leercontract, werkplekleren) heeft in Duitsland een hoge status, terwijl het in België eerder bezocht wordt door leerlingen die het op school niet meer zo zien zitten. Om het heel kortaf te zeggen: onze bevolking staat niet achter het werkplekleren en daarmee is de kans dat het hier wordt overgenomen, niet vanzelfsprekend.

Een ander voorbeeld. In de Scandinavische landen zitten leerlingen tot zestien jaar samen in dezelfde school en in dezelfde klassen. Pas dan is er een gerichte keuze voor verdere studie of voor een beroepsgerichte vorming. In die landen begrijpt men niet waarom in België en Nederland leerlingen al moeten kiezen op twaalf jaar (in Duitsland zelfs al op tien jaar). Duidelijker kan het verschil in culturele en historische achtergrond bijna niet worden geïllustreerd. Het is bijna zoals de grote filosoof Pascal het ooit in een van zijn 'pensées' stelde: waarheid in één land kan een dwaling zijn in een ander ("Vérité en-deça des Pyrenées, erreur au-delà").

De traditie van het vergelijken ligt dus in het leren van elkaar. Die traditie staat tegenover het beoordelen en rangschikken van onderwijssystemen. In dat geval meet het vergelijken zich erg grote pretenties aan. Dat is vaak ten onrechte en bovendien het resultaat van beperkte metingen. De media versterken vaak de met veel statistisch jargon omgeven zogenaamde kwaliteitsrangschikkingen. Het is realistisch het te houden bij de traditie van het leren van elkaar, rekening houdend met diverse contexten. We zien hoe men in andere landen omgaat met kinderen, die thuis niet de onderwijstaal spreken, hoe men omgaat met verschillen in de klas, hoe men leerlingen burgerzin

bijbrengt, hoe men de kwaliteit van het onderwijs nagaat enz. Maar telkens kom je na zo een bezoek – al of niet met een rapport – terug met de voeten op de Belgische/Vlaamse eigen bodem. Recepten uit het buitenland worden gesitueerd in de eigen receptuur en misschien passen bepaalde elementen daarvan in het eigen menu.

Die traditie van het leren van elkaar kreeg in de jaren 1960 een concurrent, voornamelijk afkomstig vanuit de economie. Het ging om de *Human Capital Theory*. Onderwijs, gezondheid en andere maatschappelijke organisaties of kenmerken zouden op een of andere manier als menselijk kapitaal de economie vooruithelpen. Er werden toen theorieën ontwikkeld over het onderwijs als investering in de economische groei. Dat was de start van een indrukwekkende lijst van studies naar het verband tussen onderwijs en economische groei, de zogenaamde *Rate of Return*. Men ging op zoek naar algemene wetmatigheden die objectief en wetenschappelijk controleerbaar zijn, om dergelijke verbanden op te sporen. Dat gebeurde via kwantitatieve verzameling van gegevens over economische parameters, die dan in verband werden gebracht met resultaten van het onderwijs. De meest populaire indicatoren voor economische productiviteit zijn nog steeds: het gemiddelde individuele inkomen en het bruto binnenlands product.

Dergelijke studies zijn nog steeds erg populair. Oorspronkelijk werden ze veel gebruikt binnen het kader van de ontwikkelingshulp door de UNESCO, de Wereldbank, stichtingen zoals de Carnegiestichting, de Rockefellerstichting en USAID (United States Aid for International Development). Het ging, zoals reeds gesteld, vaak om correlaties, die wel een verband geven, maar die te ruw waren om oorzakelijke verbanden eruit te distilleren. De evidente reden was, dat men de contexten niet onder controle had, zodat een inputfactor in systeem A wel goede resultaten opleverde, maar niet in systeem B.

Een voorbeeld ter illustratie. In bepaalde landen was er een verband tussen de klasgrootte met de schoolresultaten en in andere dan weer juist niet. Anders gezegd, een Ghanese onderwijzer met een klas van 60 leerlingen haalde even goede resultaten voor rekenen en lezen als een Oostenrijkse onderwijzer met 20 leerlingen. Als context kan je hierbij bedenken dat in Ghana het volgen van onderwijs zeer gegeerd is en een springplank vormt voor sociale mobiliteit, terwijl westerse kinderen daarvan niet meer wakker liggen. Het ligt voor de hand dat de Ghanese kinderen optimaal gemotiveerd zijn in vergelijking met de meer gekoesterde westerse kinderen.

Op dit ogenblik vinden we dit soort studies naar output terug bij de OESO (met de PISA-toetsen en de PIAAC-toets), bij de IEA (met de TIMSS- en de PIRLS-toetsen) en zelfs met een vergelijkend onderzoek naar de resultaten van burgerschapsopvoeding. Ondertussen is deze vergelijkende

toetsbeweging ook big business geworden. Ook privéconsortia beginnen brood erin te zien. Het machtige privéconcern Pearson, bijvoorbeeld, maakt een eigen *Pearson Global Index of Cognitive Skills and Educational Attainment*.

De expertise om toetsen te ontwerpen, is in de loop van die jaren sterk geëvolueerd. De statistiek daaromtrent heeft een hoge vlucht genomen, zodat de eenvoudige correlaties van voorheen aangevuld worden met statistische correctieformules om contexten mee in beeld te brengen. Ook meerniveauonderzoek komt nu vaak voor (de impact van diverse hiërarchische niveaus: bv, de klas, de school, de centrale overheid). Maar zoals reeds eerder vermeld en zoals verder opnieuw zal blijken: contexten zijn moeilijk in kaart te brengen en bovendien interageren ze met elkaar, zodat er steeds meer moet worden gecorrigeerd en zodat men ook meer en meer afwijkt van wat juf Fatima en meester Tom in de klas meemaken.

De resultaten van die onderzoeken zijn wel interessant om van elkaar te leren, zoals het in de tradities van het vergelijken gebruikelijk is, maar ze worden gevaarlijk en ook misleidend wanneer ze dienen om systemen te rangschikken en ze te evalueren. Er is in die zin wel een verschil tussen de TIMSS- en PIRLS-toetsen (georganiseerd door IEA) en de PISA-toetsen. De IEA is een organisatie die vanuit de onderzoekers aan de basis is gegroeid. Ze heeft de bedoeling gegevens uit internationaal onderzoek te verschaffen om de deelnemende systemen toe te laten hun eigen systeem en hun eigen doelstellingen te verbeteren. PISA gaat daarin verder en streeft in feite shockeffecten na met de bedoeling om besluiten bij de deelnemende landen te implementeren. Als klein zusje ten opzichte van PISA heeft de IEA minder impact. Hier en daar laten TIMSS en PIRLS zich soms verleiden tot het maken van rankings, niet zozeer uit principe maar vanuit de vraag van bepaalde deelnemers.

III. Een al dan niet verborgen agenda

Binnen de strenge normen van de wetenschappelijke onderzoeksgemeenschap is het gebruikelijk onderzoeken streng te bevragen op methodologie, uitgangspunten, aannames, gehanteerde theorieën, beperkingen en onderzoeksvragen. Op dat vlak is er bij de PISA-projecten wel een en ander aan te merken. Eerst en vooral gaat het om het mathematiseren van menselijk gedrag, een zware opgave voor de menswetenschappen. Die beschikken immers niet zoals de natuurwetenschappen, over universele schalen met gelijke begin- en eindpunten zoals, bijvoorbeeld, bij afmetingen, graden, frequenties, stralingen, voltages en noem maar verder op. Verder moet je geen vergrootglas hanteren om te merken dat

PISA, getrouw aan zijn organisatie, een uitgesproken economische agenda heeft. Ten slotte wil PISA onderwijssystemen veranderen volgens de besluiten, die uit hun data worden afgeleid en die dan ook universeel zouden gelden.

III.1. De betanijd

De Cambridge-econoom CHANG is verantwoordelijk voor de term 'betanijd', afgekeken van de klassieke psychoanalyse waar gesproken werd over penisnijd in verband met bepaalde psychische complexen. Hij vindt dat de economie, als typische menswetenschap, te voortvarend is bij het mathematiseren van economische processen en in het bijzonder in het voorspellen ervan. Menswetenschappen (alfawetenschappen) hebben een betanijd ten opzichte van de natuurwetenschappen en doen dus vergeefse pogingen om die na te bootsen en op die wijze status te verwerven. CHANG heeft niet veel moeite nodig om dit aan te tonen met zeer frappante voorbeelden voor de economie (CHANG, 2014). Op dezelfde wijze kan je redeneren voor andere menswetenschappen zoals psychologie, sociologie en onderwijskunde. Al die wetenschappen proberen menselijk gedrag te meten en op basis van die metingen verbeteringen aan te brengen. De onderzoeksgemeenschap binnen de menswetenschappen is zich daarvan bewust en onderwerpt mathematiseringen en metingen dan ook aan een strenge controle, meestal via peerevaluatie. Iedere onderzoeker die zichzelf respecteert, zal dan ook duidelijk wijzen op zijn uitgangspunten, de gehanteerde methodologie en mogelijke beperkingen voor veralgemening.

Bij menswetenschappen en dus ook bij het toetsen van leerprestaties is de grote moeilijkheid dat je zelf een schaal moet opbouwen en dus zelf een begin- en eindpunt moet creëren. Dat gebeurt via normen, targets of indicatoren. Dat zijn telkens aannames, die wellicht op een zekere consensus berusten, maar geen universele aanspraak kunnen maken. Om toch te objectiveren, maakt men vaak gebruik van het wiskundige model van de Gausscurve, die gewoon het relatieve standpunt inneemt dat je voor iedere groep een normaalverdeling kan construeren. Je hebt dan telkens een gemiddelde, een spreidingsmaat, en je kan dus groepen indelen in decielen, kwartielen of soms gewoon in een middelmatige groep, een zwakkere groep en een betere groep. De normaalverdeling is dus een universeel geldende wiskundige maat. Maar ze berust volledig op de groep waarop je ze construeert. De statistiek, hoe gesofisticeerd en complex zij ook is, werkt steeds op basis van aannames. Natuurlijk is die mathematisering vaak correct – het is tenslotte toegepaste wiskunde – maar de aannames blijven telkens voor discussie vatbaar.

Daar gaat het precies om wanneer je de twee volumes van *PISA 2015* doorworstelt. De massale hoeveelheid van data

wordt omgezet in een veelheid van tabellen, correlaties en – niet te vergeten – vaak suggesties tot te vlugge besluiten. Via de moderne software kan je de resultaten van de toetsen en vragenlijsten quasi oneindig gaan correleren met eender welke andere data. De verleiding om dan naar oorzaak-gevolgrelaties te gaan, wordt dan erg groot, zeker bij leken in het vak of bij mensen met bepaalde vooronderstellingen in het achterhoofd.

Nu is het zo dat het technisch handboek voor de constructie van de toetsen voldoet aan de normen die horen bij een ernstige toetsconstructie. Bij het beschrijven en interpreteren van de resultaten worden ook wel sporadisch nuanceringen aangeboden, maar op bepaalde methodologische tekorten krijg je geen antwoord. Van die nuanceringen zie je nauwelijks iets wanneer de OESO met die resultaten de boer op gaat. Gelukkig is die nuancingering wel aanwezig bij de Vlaamse contractanten, die de PISA-toetsen in Vlaanderen hebben afgenomen en geïnterpreteerd. Jammer genoeg zie je die nuanceringen onvoldoende verschijnen in de media.

III.2. Een economische agenda

Je kan het een economische organisatie, die toch de OESO is, niet kwalijk nemen dat zij de nadruk legt op economische doelen. Ook niet dat ze daarbij het onderwijs in het vizier neemt. De match tussen onderwijs en werkgelegenheid is, bijvoorbeeld, een zeer ernstig te nemen thema. Ook bekommernissen rond financiële geletterdheid en kennis van de economische processen zijn voor een dergelijke organisatie aandachtspunten.

Maar het gevaar van overaccentuering van de economische dimensie is duidelijk aanwezig. Competitie is een sleutelwoord, wat doet veronderstellen dat een bepaalde ideologie binnen de economie, wellicht met een neoliberale tint, een drijvende kracht is. Zo stelt het belangrijke paper *Beyond PISA 2015* als een gevolg van de doelstellingen van PISA: "Dit impliceert dat de maatstaf voor het succes van het onderwijs niet meer ligt in het afwegen ervan tegenover nationale normen, maar wel tegenover de best presterende onderwijssystemen in de wereld" (OECD, 2015, § 12, vertaling R.S.). De economische benadering van het onderwijs geraakt zo in 'overdrive'. Het economisch subsysteem in de maatschappij neemt dan een soort monopoliepositie in om het onderwijs als geheel in een samenleving te toetsen en te beoordelen. Het is niet toevallig dat PISA kiest voor zogenaamde competenties, die als nuttig worden beschouwd in tegenstelling met heel wat inhouden die in het onderwijs aanwezig zijn.

In een interview in 2013 drukte de topman van PISA dat als volgt uit: "Basically, the kinds of things that schools are traditionally best at are losing labour market relevancy very quickly" (TES, 21.06.2013). De 'competenties' zijn het

alternatief. Het is ook geen toeval dat die beperkt worden tot natuurwetenschappen, wiskunde en leesvaardigheid.

Dat er bepaalde verbanden bestaan tussen onderwijs en economische processen, kan niet worden ontkend. Maar, rekening houdend met de complexiteit van een land, van een economisch systeem en van een bepaalde cultuur, zijn die verbanden niet vanzelfsprekend en meestal niet veralgemeenbaar. Vaak zijn de connecties slechts contextueel aanwezig. Enkele aannames die impliciet vervat zitten in de OESO-ideologie, gaan als volgt.

Je zou kunnen stellen dat er een groot verband bestaat tussen het onderwijs en de rijkdom van een land. Dat blijkt uit de vroegere onderzoeken in ontwikkelingslanden. In die landen klopt dat ook. Het gaat dan om een nog onvoldoende ontwikkeld en veralgemeend onderwijs waardoor de weerslag daarvan op de economie vrij duidelijk is. Dat is echter niet meer zo in hoogontwikkelde landen met een sterk uitgebouwd onderwijssysteem. Je zou kunnen stellen, hoe meer jaren onderwijs een bevolking geniet, des te beter het land economisch zal vooruitgaan. Die correlaties zijn er wel in bepaalde gevallen, maar je kan ze ook omkeren: hoe rijker een land is, des te meer jaren onderwijs volgen zijn inwoners. Dergelijke kip-of-eiverbanden zijn legio in het OESO-denken. Een andere verleidelijke aanname is, dat vooral het peil van het onderwijs de beste indicator is voor een rijke economie. Dat komt dan erop neer dat je vooral moet investeren in de bollebozen, want die zullen het verschil maken. Zie, bijvoorbeeld, de klacht dat het aantal hoog scorende Vlaamse leerlingen in wiskunde gedaald is tot 21 % (dat is nog 11 % meer dan het OESO-gemiddelde). Maar een alternatieve redenering is dat je slechts een minderheid aan zeer sterke bollebozen nodig hebt om innovatie tot stand te brengen. Je hebt wel een grote massa nodig aan *subbollebozen* en uitvoerders (met een goede algemene vorming tot einde leerplicht) om die innovaties uit te voeren en ze in praktijk en in de economie om te zetten. Hoe komt het, bijvoorbeeld, dat de VS, met een erg matig onderwijs, toch de eerste economie van de wereld is? Het is erg eenvoudig: ze zijn welvarend genoeg om een beperkt aantal (betalende) elitescholen te organiseren en bovendien plukken zij de bollebozen weg uit het buitenland. Bekijk maar eens de lijst van vreemde namen bij de hoogleraren van de elite-universiteiten in de VS.

Een andere aanname is dat de Aziatische landen met hun hoge prestaties op de PISA-toetsen onze economieën zullen overtreffen en dat precies door hun onderwijs. Alternatieve onderzoeken hieromtrent wijzen eerder op de ongenadige prestatiedruk en inzet bij de bewoners van die landen en niet zozeer op het onderwijs. Die prestatiedruk, gepaard met sterke groepsdruk en volgzaamheid – en dus niet zozeer het onderwijs – is een cultureel fenomeen dat de economische prestaties sterk bepaalt.

Consequent met de aard van de economie en de onderwijskunde als menswetenschap zijn de meetpunten voor discussie vatbaar. Als je PISA-prestaties correleert met een maat als het bruto binnenlands product, vergelijk je dus op basis van die indicatoren of maatstaven. Als je andere maatstaven neemt, krijg je uiteraard andere resultaten. Bijvoorbeeld, een indicator van inkomensongelijkheid, een geluksindicator, een tevredenheidsindicator, een indicator voor corruptie, een indicator voor ondernemingszin, voor solidariteit enz. Het kenmerk van al die indicatoren is steeds dat ze op bepaalde aannames berusten en daarover is vrijwel altijd discussie mogelijk.

Verder kan je ook stellen dat economische bedrijvigheid meer en meer geglobaliseerd verloopt en de grenzen van de natiestaten en dus van de deelnemende PISA-landen overstijgt.

Het blijkt dus dat de verbanden tussen onderwijs en economische groei complex en meerdimensioneel zijn.

III.3. PISA als activist*

Uit wat voorafgaat, mag al blijken dat PISA geen neutrale speler is in de onderwijsrealiteit. In de reeds vermelde paper 'Beyond PISA 2015' lezen we dat PISA het volgende beoogt: "Voorzien in internationaal vergelijkbare evidentie over de kwaliteit, de gelijkheid en de doelmatigheid van schoolsystemen" (OECD, 2015, 2, vertaling R.S.). Dat is duidelijke taal. Het gaat niet zozeer om hulpmiddelen voor het lokale onderwijsbeleid, zoals de organisatoren van TIMSS en PIRLS beogen, maar wel om een arbitrage van de bestaande onderwijssystemen in de wereld. PISA presenteert zich echter niet alleen als diagnosticus, maar ook als beoordelaar en voor de hand liggende adviseur voor het lokale onderwijsbeleid.

De OESO is niet afkerig van een soort mediaspektakel waarbij men wereldwijd aankondigt op welke datum en welke plaats de boodschap zal worden verkondigd. Daarachter zit een geoliede machine, waardoor je, bijvoorbeeld, overal ter wereld dezelfde 'sheets' over de resultaten kan vinden. De resultaten worden gretig gepubliceerd in allerlei media, voornamelijk met economische achtergrond. Die komen dan weer terecht bij politici, die wel gretig economisch nieuws lezen, maar zelden pedagogische teksten. En zo wordt de PISA-hype een politiek feit in allerlei landen. En zo komt het ook dat economen dan hun wijsheden gaan verkondigen over wat er met het onderwijs moet gebeuren. Een vooraanstaand econoom kon op die wijze op het programma *Terzake* stellen dat het onderwijs in Vlaanderen dringend iets moest doen aan de bollebozen en dat het ook meer vrijheid aan de scholen en leraren moet geven. Dit alles zonder weet te hebben van het feit dat de vrijheid die leraren en scholen in Vlaanderen genieten, vrijwel nergens anders in de wereld zo sterk is.

Populaire voorstellingen zijn mogelijk via het systeem van ranking. Niets is zo verleidelijk als een rangschikking van voetbalploegen in competities. Iedere ploeg wil wel stijgen in die competitie en dat zullen heel wat trainers en coaches geweten hebben. De ranglijst voldoet aan die behoefte. Alleen is het eenvoudige criterium (het aantal doelpunten) niet toepasbaar op de PISA-resultaten. Het criterium komt van een geconstrueerde schaal rond zowat 500 punten. Daarbij kan je dus met veel eenheden werken. Tien punten verschil is al veel en twintig nog meer. Stel dat je een schaal zou maken met een gemiddelde van 10 punten, dan zouden 10 punten uit de 500-schaal kunnen gelijkstaan aan 0,2 punt, een score waarover niemand zich verder druk maakt.

Overigens kan je de vraag stellen hoe één cijfer per land de veelzijdige realiteit van vooral grotere landen kan weergeven. Zelfs in het kleine België zie je verschillen tussen de Nederlandstalige en Franstalige Gemeenschap. Wat moet je dan zeggen van verschillen tussen Catalonië en Andalusië in Spanje, tussen Massachusetts en Alabama in de VS, tussen de rijke provincies in China tegenover de arme enz. Alleen al het aantal inwoners van een land zou een aanduiding moeten zijn van de zinloosheid van dergelijke categorisering.

IV. Sterke contextgevoeligheid van de data

Het is uitermate moeilijk om binnen de menswetenschappen universele wetmatigheden te vinden zoals dat het geval is bij de natuurwetenschappen. Dat heeft niet alleen te maken met de onvoorspelbaarheid van het menselijk gedrag, zowel individueel als in groepen. Het heeft ook te maken met de grote diversiteit aan contexten waarin het geteste gedrag plaatsvindt. Vandaar dat je in vrijwel alle onderzoeken besluiten krijgt in de zin van: 'onder deze omstandigheden' en 'binnen de gegeven context'. In bepaalde gevallen kunnen een heleboel onderzoeken (meta-analyses) uitmonden in een trend, die een vrij grote waarschijnlijkheid geeft. Dat is wat auteurs als HATTIE hebben beoogd bij het zoeken naar trends bij het toepassen van didactische en pedagogische maatregelen in het onderwijs (HATTIE, 2012, BELLENS en DE FRAINE, 2012). Maar zelfs dan kan je die tendensen niet als een receptenboek – klaar voor gebruik – hanteren.

Bij de toetsen van PISA speelt die context sterk mee, vooral bij de keuze van de inhoud van de vragen, enerzijds, en bij de culturele omgeving van de schoolsystemen, anderzijds.

IV.1. Inhoudelijke context

Ook al pretendeert PISA vragen te formuleren, los van de programma's, toch beïnvloedt de manier waarop bepaalde

leerinhouden en vakken in de verschillende systemen geprogrammeerd zijn, de mogelijkheden om de vragen op te lossen. De leerplannen in de diverse systemen verschillen en bovendien kunnen de leerplannen vroeger of later dan 15 jaar (de leeftijd van de afnames) gepland zijn. Zo zijn er landen waar men al op 15 jaar heeft kennism gemaakt met differentiaal rekenen en in andere landen dan weer niet. Daarnaast is het vrij logisch dat de verschillende systemen ook verschillende klemtonen kunnen leggen in de keuze van inhouden. Als je in tropische landen leeft, zal de manier waarop je met biologie, klimaat en energie omgaat anders zijn dan in koudere landen. Zo is ook bekend dat grammatica in het Franse taalonderwijs in de lagere school vaak de helft van de tijd inneemt, terwijl dat aandeel in Engeland minimaal is. Het is wel zo dat binnen PISA-vragen die cultureel 'biased' zijn, in bepaalde landelijke versies worden verwijderd. Maar daardoor heb je ook geen zicht op de verschillen die in dat systeem aanwezig zijn. Daarnaast kunnen de verschillende culturen ook andere klemtonen leggen in het beheersen van de leerstof. In bepaalde landen hecht men meer belang aan bepaalde vakken en inhouden dan in andere systemen. Dat kan, bijvoorbeeld, al blijken uit het aantal uren in het lessenrooster. Het is, bijvoorbeeld, ook goed zichtbaar dat de vragen voor wiskunde bij de TIMSS-afnames meer relevant zijn voor hogere studies in technologische sectoren, wat bij PISA minder het geval is.

Zelfs de structuur van het onderwijs bepaalt in grote mate de achtergrond om de vragen van PISA te beantwoorden. Kijkend naar de Vlaamse resultaten voor wetenschappelijke geletterdheid in 'PISA 2015', zien we zwakke prestaties voor tso en bso. Eigenlijk is dat volkomen logisch omdat in België het technisch- en beroepsonderwijs voluit gaat na de eerste graad (zelfs in bepaalde scholen al vroeger). Dan wordt algemene wetenschappelijke geletterdheid vervangen door toegepaste technische inhouden op gebied van mechanica, elektriciteit, landbouw, tuinbouw en dergelijke. In vrijwel alle systemen in de wereld begint men pas met die toepassingen na 15 jaar, zodat daar algemene wetenschappelijke geletterdheid langer wordt onderwezen.

IV.2. Culturele context

In de comparatieve pedagogiek is er de algemene basisregel, dat het vergelijken van onderwijssystemen niet mogelijk is zonder de contexten in rekening te brengen. Dat is overigens erg moeilijk en het vraagt vaak een grondige studie van de geschiedenis en de cultuur van een land. Het is daarom ook vrijwel onmogelijk om onderwijssystemen op elkaar te laten gelijken. Vorming en onderwijs zijn nu eenmaal zeer waardegebonden. In landen met erg centralistische systemen kan men die waardeverschillen beperken. Waardevorming wordt dan meer aan de private sfeer toegewezen. Centralistische systemen stellen zich ook vrij cognitief op omwille van een veronderstelde neutraliteit. Op die wijze zijn dergelijke

systemen ook meer vatbaar voor centrale testsystemen die uiteraard het cognitieve gedrag meten. In landen waar meer decentralisering is, laat men ook meer verschillen toe en wordt plaats gelaten voor waardevorming. Die is dan uiteraard minder vatbaar voor objectieve toetspraktijken.

Maar zelfs in centralistische systemen zit het onderwijs ingebed in een sociologische, psychologische en demografische omgeving. Het systeem onderwijs krijgt daardoor een cultuurbepaaldheid, die het onderwijs daaraan onderschikt. Dat gaat zo ver dat je in een aantal gevallen de vraag kan stellen of de resultaten van de toetsen wel te wijten zijn aan het onderwijssysteem.

Neem de hoog PISA-presterende Aziatische landen. Ze zijn stuk voor stuk gekenmerkt door een uitzonderlijk zware prestatiedruk, die afkomstig is van de ouders en dus in feite van de culturele omgeving. Centraal staan in die landen de afsluitende toetsen rond 17-18 jaar die uitermate bepalend zijn voor de toekomst van de jongeren. Falen in die landen is een blamage voor de ouders. De centrale afsluitende toetsen bepalen de hele schoolcarrière van de kinderen. De druk daarop daalt neer naar de lagere jaren tot zelfs naar het basisonderwijs. Het gevolg daarvan is dat al vroeg gegrepen wordt naar het alomtegenwoordige privéonderwijs. Kinderen en jongeren gaan na schooltijd massaal naar (betalend) 'shadow'-onderwijs. Die situatie wordt in westerse landen met verbazing bekeken. De termen daarvoor geraken ingeburgerd: 'gaokao' in China, 'hangwo' in Korea en 'yuku' in Japan. Naarmate de jaren stijgen, betekent het dat jongeren tot 22u in de avond naar de privéschool gaan. Heel de jeugdperiode is gecentreerd rond dat eindpunt van het nationale examen. De gevolgen voor het dagelijkse leven van de jongeren zijn in onze westerse ogen onbegrijpelijk. Het gaat om een eeuwenoude confucianistische prestatiecultuur waarbij bepaalde intellectuele prestaties de status van een individu bepalen. Groepsdruk, gehoorzaamheid, volgzaamheid, weinig interactie zijn kenmerken van dit soort onderwijs. Dat die landen hoog scoren, is dus niet te wijten aan de school, want die wordt nauwelijks ernstig genomen in deze 'shadowstructuur'.

Het is dan ook merkwaardig te zien hoe officiële landelijke delegaties de heilige graal van het succes in die landen gaan zoeken. In een revelerende documentaire heeft de Vlaamse cineast Steven DHOEDT samen met een Zuid-Koreaanse collega een documentaire gewijd aan deze prestatieobsessie in Zuid-Korea (DHOEDT, 2016). De film is representatief voor alle Aziatische prijswinnaars bij de PISA-toetsen. Daarbij stelt de Amerikaanse hoogleraar ZHAO, die zelf zijn onderwijs in China heeft genoten en auteur is van een spraakmakend boek over de *big bad dragon* het volgende: "Het Chinese onderwijs verstrikt creativiteit, beperkt nieuwsgierigheid, onderdrukt individualiteit, schaadt de gezondheid van de kinderen,

bezorgt kinderen en ouders stress, zet leraren en directies aan tot corrupte toestanden en vereeuwigd sociale onrechtvaardigheid en ongelijkheid" (ZHAO, 2014, 187, vertaling R.S.).

Een interessante studie van KAMENS vergelijkt de resultaten in 2012 van drie hoog scorende systemen voor wiskunde: Finland, Shanghai en Vlaanderen. De resultaten van de drie systemen liggen in dezelfde orde, maar de systemen en de onderliggende culturen verschillen grondig. Zelfs de observaties in de klas wijzen op grote verschillen: Vlaanderen is didactisch gezien het meest probleemoplossend in de klas, Finland past eerder traditionele frontale methodes toe en Shanghai bereikt de resultaten door uitvoerig drillen en herhalen via het onderwijs, maar nog meer via het 'shadowonderwijs'. Het is duidelijk dat ondanks analoge prestaties in de toetsen "these larger contexts matter" (KAMENS, 2013, 132).

Een andere studie vergelijkt via klasobservatie de aanpak van wiskunde bij de hoog PISA-scorende systemen Finland en Vlaanderen in 2009. Het besluit is dat het succes van de twee systemen toe te wijzen is aan een verschillende karakteristiek in de twee respectieve culturen. Bijvoorbeeld, een zeer diepgewortelde leescultuur in Finland tegenover een meer probleemoplossende aanpak in Vlaanderen (ANDREWS, 2014).

Misschien minder uitgesproken, maar toch duidelijk zijn de cultuurverschillen bij de westerse landen. Hoe kijkt men in de verschillende landen aan tegen het onderwijs? Hoe divers of homogeen is de bevolking in een bepaald land? Is er competitie tussen scholen bij het rekruteren van leerlingen? Welke status heeft een leraar? Is er kinderopvang en voorschools onderwijs? Zijn er opvallende demografische en aardrijkskundige kenmerken? Wordt een systeem centralistisch-bureaucratisch geleid of is er veel autonomie voor scholen? Enzovoort. Als je al die factoren en onderfactoren rubriceert, kom je tot een indrukwekkende lijst van interveniërende factoren, zowel op het niveau van leerlingkenmerken als op het schoolniveau. Maar ook op het niveau van de leraren, de gezinsstructuren en opvoedingsgebruiken daarbinnen. Ten slotte overheersen diep ingebedde waarden en normen in een bepaalde samenleving. Al die factoren kruisen met elkaar en zo krijg je zeer complexe contextuele situaties, die het vergelijken en zeker het beoordelen alsmoar moeilijker maken.

Misschien moet je hier het gezond verstand laten spreken. Een soort universeel en cultuuronafhankelijk *curriculum* dat door de OESO en zijn PISA-experts wordt gepromoot, is een illusie. Je kan onmogelijk stellen dat jongeren van 15 jaar dezelfde ambities, voorkeuren, problemen en uitdagingen hebben in, bijvoorbeeld, Mexico, Griekenland, Denemarken, Vietnam of Texas. Ze zullen dus ook eigen vaardigheden en competenties daartoe ontplooien.

V. Methodologische discussies

Het laat zich raden dat, gezien de moeilijke situatie van de menswetenschappen bij het meten van menselijk gedrag, allerlei methodologische problemen zullen opduiken. Dat is niet anders voor een zo ambitieus project als PISA. Om niet al te technisch te worden, komen hierbij de methodologische kwesties erg beknopt naar voor. In wetenschappelijke kringen worden ze uitvoerig besproken.

Een eerste element is de steekproeftrekking. Het gaat om de 15-jarigen in een bepaald land of een bepaalde regio (bv. vier regio's in China). Bij die 15-jarigen zitten, bijvoorbeeld, ook zittenblijvers; er zijn landen met veel en met weinig zittenblijvers. Ook stelt zich de vraag in welke mate kinderen met een beperking in de steekproef zitten. Waar ligt de grens van inclusie? Verder heb je ook het feit dat in een aantal landen heel wat leerlingen uitvallen vóór de leeftijd van 15 jaar. In een aantal landen is ook de kinderarbeid mee verantwoordelijk voor onregelmatig schoolbezoek of voortijdig afhaken. Zo kom je uit bij een geprivilegieerde groep. In Shanghai en andere Chinese regio's worden de kinderen van talrijke niet-geregistreerde inwoners (afkomstig uit andere Chinese regio's) niet meegeteld. Dat zou bijna hetzelfde aantal zijn als de deelnemende jongeren. Verder krijg je de impact van subculturen. Zo blijken Chinese migranten in Australië ver boven het niveau van de Australiërs te scoren. Hun scores zijn vergelijkbaar met de topscoorders van Shanghai. Daarentegen blijkt Australië als geheel, met die topscoorders erbij, op de 19de plaats te scoren.

Een ander probleem heeft te maken met controle op de vertalingen. De taalstructuur van, bijvoorbeeld, een Germaanse taal tegenover het Fins, het Chinees of het Arabisch is zo verschillend dat het vertalen van oorspronkelijk Engelse teksten en items naar die talen niet zo vanzelfsprekend is. Het is niet uitgesloten dat er op die wijze ook weer ruis ontstaat in de gelijkheid van de vraagstellingen. Dat geldt zeker voor de vragen in verband met leesvaardigheid. Lezen in het Chinees, een typische beeldtaal, veronderstelt, bijvoorbeeld, andere onderliggende intellectuele disposities dan lezen in een westerse taal.

Verder is het niet vanzelfsprekend dat de gestelde items in alle systemen dezelfde moeilijkheidsgraad hebben. Nochtans is dat een voorwaarde voor de toepassing van het gebruikte schalingsmodel. Het gebruikte zuivere Rasch-model onderstelt dezelfde moeilijkheidsgraad van de vragen in alle landen en dat is precies een fundamentele kritiek op de PISA-toetsen.

Zoals dat bij surveys vaak het geval is, worden niet alle vragen aan alle testpersonen aangeboden. De duur van de test zou dan zodanig toenemen dat de toetsafname

niet onder de beste omstandigheden kan gebeuren. Er zijn, bijvoorbeeld, 81 vragen voor wiskunde en 184 voor wetenschappelijke geletterdheid. Daarom krijgen verschillende leerlingen verschillende delen van de toets aangeboden. Via een statistisch schattingsmodel wordt de verkregen score geëxtrapoleerd naar de hele toets. Ook die benadering geeft ruis op de resultaten.

Heel wat items van de aangeboden competenties worden in een situatie beschreven. Ze bevatten een verbale leescomponent, die ook weer, naargelang van de leesvaardigheid in bepaalde landen, verschillen kan veroorzaken. Daarenboven gaat het vrijwel altijd om meerkeuzevragen. Daarbij geldt dat, als onjuist bestempelde alternatieven, door creatieve en verstandige leerlingen vaak als zinvol worden gekozen.

Algemeen gesproken, kan je ook de problematiek van momentopnames tegenover een continue cohortemeting stellen. Ook kan je vragen stellen over de validiteit van het al dan niet vooruit- of achteruitgaan in de rangschikking over zovele jaren. Als er, bijvoorbeeld, meer ontwikkelingslanden zullen deelnemen, kan je zo voorspellen dat het gemiddelde van het geheel zal dalen.

In 58 van de 72 deelnemende systemen werden de toetsen via computer aangeboden. Hier komt dan weer een nieuwe factor naar voren. Naarmate de testpersonen meer beslagen zijn in ICT-vaardigheden, bouwen ze een voorsprong op in de antwoorden. In replicatieonderzoek in Duitsland door Köller van de universiteit van Kiel, werd als replicatie een toets in de geschreven en in de ICT-versie aangeboden. Daaruit bleek dat de factor ICT-vaardigheid de uitslag tot 20 punten verschil beïnvloedde (DREBES en VOLLMER, 2016).

Je kan ook vragen stellen bij de mogelijke interpretaties op basis van de vragenlijsten. Het gaat dan om factoren van schoolklimaat en zelfs methodes op de klasvloer die men verbindt aan de resultaten op de toetsen. Het gaat om percepties, die niet noodzakelijk de realiteit weerspiegelen. Bovendien zijn de gebruikte benamingen van methodieken in de klas erg onzuiver en te algemeen om tot relevante besluiten voor de klasvloer te leiden. Een positieve associatie van leraarsgerichte didactiek met goede resultaten voor wetenschappen, staat zo in schril contrast met een negatieve associatie met onderzoekende methodiek (bv. laboproeven). Vanuit de wetenschappelijke community zal men de vaagheid van de begrippen aanklaarten en daarnaast stellen dat je klasobservaties nodig hebt om de impact van en de methode, gelinkt aan concrete contexten, te bepalen.

Wanneer het om een dergelijk ontzaglijke onderneming gaat met contractanten in 72 verschillende landen of regio's, is de controle op de verzameling en inventarisatie van de talrijke data een haast onmogelijke onderneming.

Dat is op zichzelf al moeilijk bij testsituaties binnen één land. Wanneer je dat in 72 regio's moet waarmaken, krijg je een moeilijkheidsfactor in het kwadraat. De Duitse expert BLASIUS en zijn Canadese collega THIESSEN hebben meticuleus onderzocht hoe dit voor een onderdeel van *PISA 2009* gebeurde in vrijwel alle deelnemende systemen. Ze deden dat aan de hand van het onderdeel vragenlijst voor directies. Slechts 16 van de deelnemende landen konden de toets van kritiek doorstaan. Tien van de deelnemende landen werden bestempeld als volledig onbetrouwbaar. Onder meer in honderden kopijen van vragenlijsten bleek dat directies gewoon steeds dezelfde blokjes van de alternatieven (bij de meerkeuzevragen) hadden ingevuld (BLASIUS en THIESSEN, 2015). Dit laatste voorbeeld verwijst dan ook weer naar een algemene opmerking, dat de meeste antwoorden via meerkeuzealternatieven worden gegeven.

Gezien het ambitieuze karakter van de PISA-toetsen om te diagnosticeren, te oordelen en te adviseren op basis van een uitermate grote verzameling van gegevens in 72 landen of regio's, kan die onderneming als een methodologisch bijna ondoordringbaar woud worden beschouwd. Als je de strengheid bekijkt waarmee allerlei kleinschaligere onderzoeken binnen de wetenschappelijke onderzoeksgemeenschap worden beoordeeld, kan de vraag worden gesteld hoe dit megaproject erin slaagt om daaraan grotendeels te ontsnappen. De nuanceringen die de onderzoekers zelf aanbrengen en ook de correctheid van de technische handleiding bij het project, volstaan daarom niet. De onvolkomenheden en zwakke aspecten zitten verscholen in lijvige boekwerken die vrijwel door niemand grondig worden gelezen. In feite gaat het erom dat via eigen samenvattingen de bakker zijn eigen gebak beoordeelt. In elk geval zou de publieke bekendmaking en de rapportering van de resultaten een flink stuk bescheidener kunnen zijn. Dat is gebruikelijk bij wetenschappelijk onderzoek en zeker bij het onderzoek binnen menswetenschappen.

VI. Een verbazingwekkende impact

Het is verbazingwekkend welke impact de driejaarlijkse PISA-toetsen hebben in vele landen. Het voorbeeld van Duitsland is erg illustratief voor die impact. Het begon met de befaamde *PISA-shock* in Duitsland in 2001 na de afname van de eerste PISA-toetsen. Het waren paniecreacties die leidden tot allerlei maatregelen, onder meer het oprichten van centrale examens en van centrale toetsinstituten. PISA werd een politieke hoogvlieger en de PISA-ranking werd als norm gesteld voor de kwaliteit van het onderwijssysteem. Sindsdien wordt elke PISA-presentatie in Duitsland een hype in de media en de politiek. Met een vergrootglas wordt naar de rankings en de punten gekeken en politici bouwen daarop allerlei verhalen, uiteraard verschillend

naargelang de regerende coalitie of de oppositie. Bepaalde maatregelen worden dan gezien als al dan niet bevorderlijk voor de vooruitgang of achteruitgang in de ranking. Bijvoorbeeld, het uitbreiden van het onderwijs naar de namiddag (Ganztagsschulen) of het verkorten van het Gymnasium van 9 naar 8 jaar¹. Veel levert het allemaal niet op. Duitsland ging na 2012 achteruit voor wetenschappen en leesvaardigheid.

Het voorbeeld van Duitsland geldt, *mutatis mutandis*, voor heel wat landen. Nog enkele voorbeelden. De regering van Wales stelde als norm om voor de resultaten bij de eerste 20 in de ranking te geraken. Niets werd daartoe onverlet gelaten. Programma's werden aangepast aan de PISA-toetsen en oefeningen in het beantwoorden van PISA-toetsen werden ingevoerd. Jammer genoeg voor Wales baatte het niet. Tegenover 2012 ging het achteruit in de rankings voor wetenschappen en leesvaardigheden. Alleen voor wiskunde bleef het resultaat constant. Op dezelfde wijze reageerden landen als Spanje, Zwitserland en Mexico. In enkele deelstaten van Duitsland werd het voorbeeld van Wales gevolgd, maar dus ook zonder echt gunstig gevolg.

De rankings leiden zelfs tot een soort bedevaartcultuur. Finland was tot voor kort het geliefkoosde model. Het werd onder de voet gelopen door delegaties uit allerlei landen die daar gingen opsnuffen hoe je zulke kwaliteit kan bekomen.

In 2009 kwam Polen als de nieuwe 'place to be' naar voren. Polen werd geprezen voor zijn grote vooruitgang omdat het vanaf 2000 tot 2009 opklom van erg ver naar de 15de plaats voor leesvaardigheid; in PISA-punten van 479 tot 500, zijnde de grootste groei. Ook daar werd weer gekeken naar wat de Polen allemaal verwezenlijkt hadden in hun onderwijsbeleid. De grote stijging van het aantal studenten in het hoger onderwijs aldaar zou tot voorbeeld moeten strekken. Ondertussen is Polen sinds 2012 in scores naar beneden gegaan: van de 9de naar de 22ste plaats voor wetenschappen, van 17 naar 14 voor wiskunde en van 10 naar 13 voor lezen.

Shanghai houdt stand, maar met de drie nieuwe regio's in China (eveneens rijke regio's) zie je een dalende tendens in de punten en rankings.

Tijdens een symposium georganiseerd door de Vlaamse organisatie voor ontwikkelingshulp in het onderwijs VVOB, werd uitvoerig ingegaan op de resultaten van Vietnam op de PISA-toetsen in 2012. Als land in ontwikkeling scoorde Vietnam in 2012 erg hoog; in 2015 bleef het constant voor wetenschappen (plaats 8), maar ging het achteruit voor wiskunde (van 17 naar 22) en lezen (van 20 naar 32). Onderzoekers die deze casus grondig bestudeerden, wezen op dezelfde kenmerken van cultuur, die hierboven al werden geschetst als de confucianistische prestatiedruk en daarbij ook nog sterke drillpraktijken. Ook waren de resultaten wellicht overschat doordat van de leerlingen die

op 12 jaar nog op school zaten, er 23,40 % afhaakten tussen 12 en 15 jaar (ROLLESTONE en KRUTIKOVATO, 2014).

In dit verband is de opmerking van de Harvard-hoogleraar KORETZ relevant. Hij ziet op basis van onderzoek over een veelheid van externe toetsen, een soort zaagtandprofiel ('sawtooth') ontstaan in de resultaten. Bij een nieuwe toets zie je vaak een resultaat dat vrij bescheiden is. De volgende keer is er een stijging, maar na verloop van tijd herstelt het resultaat zich naar de beginsituatie (KORETZ, 2010).

De nieuwe bestemming is nu Estland. Totdat ook daar de zaagtand van KORETZ zijn werk zal doen.

De impact gaat zo ver dat de categorieën 1 en 2 van de PISA-resultaten in allerlei landen als norm worden genomen voor een minimum aan geletterdheid. Nu is het concept geletterdheid erg bediscussieerbaar en het is bovendien sterk cultuurgebonden. Ook de Vlaamse Regering gaat daarin mee. Maar wat is geletterdheid? De meningen verschillen daarover sterk. Niet zelden is het een intellectuele elite die bepaalt wat iedere mens minimaal moet beheersen. Moet je een werk van Hugo Claus kunnen opnoemen? Moet je de spelling van de werkwoorden correct beheersen? Moet je de staatsinrichting van België met bevoegdheden van gemeenschappen en gewesten kunnen verduidelijken? Moet je het openbaar vervoer kunnen gebruiken en hoe? Enzovoort.

Deze illustraties maken duidelijk dat dit soort obsessie voor de PISA-rankings, rekening houdend met de methodologische twijfels en de culturele contexten, geen zin heeft. De fascinatie voor die rankings is niet terecht. De verzameling van de data en de validiteit ervan laten dergelijke ingrijpende conclusies niet toe. In de internationale reacties erop zie je telkens mechanismen van politieke 'cherry picking'. De resultaten worden vaak gebruikt als smeermiddel voor de politieke besluitvorming. Terwijl het helemaal niet vanzelfsprekend is hoe een bepaalde maatregel leidt tot een beter onderwijs, in dit geval beperkt tot een beter resultaat op de PISA-toetsen.

De sluipende normering die uitgaat van deze PISA-resultaten, baart zorgen. De OESO profileert zich meer en meer als een 'global education governance'. De ambities reiken steeds verder, zodat ook nieuwe toetsen in het vooruitzicht worden gesteld op gebied van vreemde talen, burgerzin, creativiteit en de zogenaamde competenties voor de 21ste eeuw. PISA stelt, bijvoorbeeld, als optie ook PISA-toetsen voor scholen ter beschikking: de *PISA-based Test for Schools*. De ambities van universele geldigheid en meetbaarheid worden door de politieke weerslag in een aantal landen nog versterkt. Politici gaan immers gemakkelijk mee in economische discours. Via het krachtige netwerk van de OESO komen de PISA-resultaten uitvoerig aan bod in allerlei economisch geïntereerde publicaties. Het is overigens ook zo dat politici meestal geen prioriteit

geven aan de lectuur van pedagogisch geïnspireerde teksten. Er zijn ook slechts weinig gegadigden om de uitgebreide en complexe rapportering grondig door te nemen. Meestal gaat het om samenvattingen of zelfs samenvattingen van samenvattingen, al dan niet gekleurd volgens de bron.

De verborgen normering via de claims van universaliteit staat in schril contrast met de discussies over de grenzen van de globalisering tegenover de eigen culturele samenleving. Terwijl de globalisering meer en meer een feit wordt op het gebied van ecologie, verkeer, veiligheid en economie, wordt de behoefte aan een vorm van culturele eigenheid duidelijker en meer geprononceerd. En het onderwijssysteem is precies de belangrijkste motor om die culturele eigenheid (historisch, sociologisch, geografisch, demografisch) te bewerkstelligen. Onderwijssystemen zullen daarom altijd verschillen.

Het zou overigens een goede oefening kunnen zijn om de prestaties van hoog presterende landen af te zetten tegenover de problematieken die in die landen overheersen. Denk, bijvoorbeeld, aan de kloof tussen arm en rijk, de milieuverontreiniging, het psychisch welbevinden (bv, suïcidecijfers), vrijheid van meningsuiting, het rechtssysteem (doodstraf bv), de behandeling van minderheden enz. In westerse landen zijn *curricula* erg bekommerd voor die belangrijke ethische elementen in een samenleving. Maar die meet je niet via een economisch discours.

VII. Conclusie

De kritische vragen zijn gesteld. Als stelling geldt dat de ambities van de PISA-toetsen overgeaccentueerd worden en dat de resultaten een betekenis krijgen die de validiteit van de data en de dataverzameling ver te boven gaat.

Het betekent niet dat de reusachtige en dure inspanningen, die het project PISA kenmerken, geen waarde hebben. PISA verzamelt een indrukwekkende hoeveelheid van data, die bovendien openbaar ter beschikking staan. Zelfs met inachtneming van de methodologische en principiële bezwaren, kunnen de data, trends en bepaalde processen detecteren. Ook kunnen ze de hypothesevorming voor verder onderzoek gunstig beïnvloeden. Het verschil tussen de sterke en zwakke presteerders in Vlaanderen is, bijvoorbeeld, een interessant signaal voor nader onderzoek. Dat geldt ook voor de kloof tussen autochtone leerlingen en die met een migratieachtergrond. Ze vormen een belangrijk startpunt voor reflectie in de deelnemende landen, zonder dat ze een receptuur worden. Het wordt dan toegepast onderzoek met het afwegen van de eigen context. De data van PISA zijn dus niet het probleem, maar wel de manier waarop ze worden gebruikt.

Uiteindelijk blijft het eindoordeel over de kwaliteit van een systeem de mate waarin het tegemoetkomt aan de doelstellingen die in een bepaald land worden vooropgesteld. En die kunnen in grote mate anders zijn dan in andere landen of systemen.

Je kan een onderwijssysteem niet samenvatten in enkele cijfers of punten. Daarvoor is het te complex en dus heeft het ook geen zin om rankings te maken of ten minste die rankings te promoten. Misschien is het best de wet van de econoom Charles GOODHART voor ogen te houden. GOODHARTs law zegt het als volgt: "When a measure becomes a target, it ceases to be a good measure".

Literatuur

- ANDREWS, P. (2014), "What does PISA performance tell us about mathematics teaching quality. Case studies from Finland and Flanders" in H.D. MEYER en A. BENAVIDOT, *o.c.*, 99-116.
- BELLENS, K. en DE FRAINE, B. (2012), *Wat werkt? Kenmerken van effectief basisonderwijs*, Leuven, Acco.
- BLASIUS, J. en THIESEN, V. (2015), "Assessing response simplification and data fabrication", *Social Science Research*, 52, 480-493.
- CHANG, H.-J. (2014), *Economie. De gebruiksaanwijzing*, Amsterdam, Nieuw Amsterdam Uitgevers.
- CHEN, X., HASTINGS, P.D., RUBIN, K.H. *et al.* (1988), "Child-rearing attitudes and behavioral inhibition in Chinese and Canadian Toddlers", *Developmental Psychology*, 34, 677-686.
- DHOEDT, S., *Reach for the sky* (documentaire), www.visualantics.net.
- DREBES, J. en VOLLMER, F. (2016), *Pisa wird zum Problem*, RP Online, 7 december 2016, www.rp-online.de/politik/pisa-wird-zum-problem-aid-1.6446861.
- KAMENS, D.H. (2013), "Globalization and the emergence of an audit culture" in H.D. MEYER en A. BENAVIDOT, *o.c.*, 117-139.
- HATTIE, J. (2012), *Visible learning for teachers. Maximizing impact on learning*, Londen-New York, Routledge.
- KLEES, S.J. (2016), "Human Capital and Rates of Return: brilliant ideas or ideological dead ends", *Comparative Education Review*, (60), nr. 4, 644-672.
- KORETZ, D. (2010), *Implications of current policy for educational measurement*, Princeton, Education Testing Service.
- LIU, J. (2016), "Student achievement and PISA rankings: policy effects or cultural explanations?" in W.C. SMITH (ed.), *The global testing culture. Shaping education policy, perceptions and practice*, Oxford, Symposium Books, 85-101.
- MA XIN JONG, C. en YUAN, J. (2013), "Exploring reasons for the success in PISA" in H.O. MEYER en A. BENAVIDOT, *o.c.*, 225-245.
- MEYER, H.O. en BENAVIDOT, A. (eds.) (2013), *PISA, power and policy. The emergence of global educational governance*, Oxford, Symposium Books, 335 p.
- OECD, PISA 2015 Results, Vol. I. (2016), *Excellence and equity in education*, Parijs, OECD, 492 p.
- OECD, PISA 2015 Results, Vol. II. (2016), *Policies and practices for successful schools*, Parijs, OECD, 468 p.
- OECD, *Beyond PISA 2015: a longer-term strategy of PISA*, Paper (2015), Parijs, OECD, 14 p.
- ROLLESTONE, C. en KRUTIKOVA, S. (2014), "Equalising opportunity? School quality and home disadvantage in Vietnam", *Oxford Review of Education*, 40, nr. 1, 112-131.
- SMITH, W.C. (Ed.) (2016), *The global testing culture shaping education policy, perceptions and practice*, Oxford, Symposium Books, 302 p.
- STANDAERT, R. (2009), *Globalisering van het onderwijs in contexten*, Leuven, Acco, 249 p.
- TES (*Times Education Supplement*), Diverse aanhalingen met aanduiding van de datum in de tekst.
- ZHAO, Y. (2014), *Who's afraid of the big bad dragon? Why China has the best (or worst) education system in the world*, San Francisco, CA, Jossey-Bass, 254 p.

