

De betekenis van internationale indicatoren voor de verbetering van het onderwijs

Lex Borghans
Universiteit Maastricht,

In opdracht van het ministerie van OCW voor de Kenniskamer OCW februari 2007.

1 Inleiding

De afgelopen jaren wordt in toenemende mate het belang ingezien van het verzamelen van statistisch materiaal om de kwaliteit van het onderwijs te monitoren. Zowel op nationaal niveau als op internationaal niveau worden stappen gezet om een beter beeld van het onderwijs te krijgen. Met name de OECD is actief bij het opstellen van indicatoren waarmee het onderwijs tussen landen vergeleken kan worden. Country reviews vullen deze indicatoren aan met meer gedetailleerde kwalitatieve informatie. Daarnaast zijn er een aantal initiatieven genomen om het kennispeil van zowel leerlingen in het onderwijs als de bevolking als geheel te meten, zodat er goed vergelijkbare indicatoren ter beschikking komen met betrekking tot de kwaliteit van het onderwijs.

Ook in Nederland is de afgelopen jaren door een aantal toonaangevende instanties nadrukkelijk gewezen op het belang van “evidence based” onderwijsbeleid (Onderwijsraad, 2006 en Cornet et al., 2006). In het verleden zijn veel veranderingen in het onderwijs doorgevoerd zonder dat er een gericht en betrouwbaar statistisch onderzoek had plaatsgevonden van de effecten die van deze maatregelen verwacht kunnen worden. Inmiddels heeft de EU ook haar ambities op het gebied van onderwijs duidelijk geformuleerd. Om een sterke positie te blijven houden in de internationale concurrentie, willen de lidstaten van de EU met hun onderwijs tot de top van de wereld (blijven) horen. Het besef dringt steeds meer door dat als we het onderwijs systematisch willen verbeteren, goede data nodig zijn om aan dit proces richting te kunnen geven.

Internationaal vergelijkbare indicatoren lijken een goede basis te vorming voor gerichte aanpassingen van het onderwijs. Door landen onderling te vergelijken kan gekeken worden welk land op een bepaald terrein het beste presteert en kunnen andere landen van de aanpak van deze “best practice” leren. Toch is het nog niet eenvoudig om bijvoorbeeld op basis van het toonaangevende OECD-rapport “Education at a Glance” (OECD, 2006) duidelijke conclusies te trekken over de tekortkomingen in het Nederlandse onderwijs en mogelijk oplossingen hiervoor. De resultaten van de PISA-studie (OECD, 2001) waarin 15-jarigen in een groot aantal landen zijn getest op leesvaardigheid, wiskunde en natuur en techniek, leidde in Duitsland tot een schrikreactie. Vrij snel werd besloten om in te grijpen en meer te investeren in het Duitse onderwijs. In de Verenigde Staten leidde vergelijkbare cijfers vooral tot ongeloof en relativering. De grondlegger van de theorie van het menselijk kapitaal Gary Becker verwoorde onlangs deze twijfel in een aantal overwegingen in een kolom op zijn website¹. Ten eerste vraagt hij zich af hoe het kan zijn dat de VS wel een hoge productiviteit per capita hebben als hun menselijk kapitaal zo slecht zou zijn. Ten tweede merkt hij op dat in het schoolsysteem van de VS 15-jarige het wellicht slechter doen, maar dat deze leerlingen deze achterstand weer inhalen als ze op de universiteit zitten. Hun Franse en Japanse medestudenten zouden tegen die tijd al te uitgeput zijn van de middelbare school om op de universiteit nog goed te kunnen presteren. Zijn derde overweging is dat scholen in de VS zich minder concentreren op academische vakken, en meer de nadruk leggen op praktische kennis. Zijn boodschap is

¹ Gary Becker, 2006, Test Scores and Economic Performance (http://www.becker-posner-blog.com/archives/2006/09/test_scores_and.html).

dus: Misschien meten we de verkeerde vormen van kennis op het verkeerde moment. Het enige resultaat dat echt telt is het eindresultaat.

Hoewel deze beschouwing een aantal fundamentele problemen van internationale vergelijkingen naar voren brengt, is de relativiserende toon niet constructief. Als men het onderwijs echt wil verbeteren heeft het niet veel zin om bij voorbaat te veronderstellen dat het eigen systeem het beste is, maar moet grondig en kritisch gekeken worden waar verbeteringen mogelijk zijn. Uit alle landen van de OECD komen signalen dat het onderwijs niet optimaal is en dat er verbeteringen mogelijk moeten zijn. Hoxby (2004) vergelijkt voor de VS bijvoorbeeld de ontwikkelingen in kosten van het onderwijs met de prestaties van de leerlingen, waaruit blijkt dat er steeds meer geld naar het onderwijs gaat zonder dat dit een hogere opbrengst oplevert. Zelfs een vergelijking met de internationale top geeft dus mogelijk nog een onderschatting van de verbeteringen die haalbaar zouden moeten zijn.

In deze notitie ga ik in op de internationale vergelijkingen van het onderwijs die de afgelopen tijd zijn uitgevoerd en bespreek ik welke betekenis deze gegevens hebben gehad voor de analyses die Nederlandse instanties van het Nederlandse onderwijs hebben gemaakt. Daarnaast ga ik in op de vraag wat de betekenis is van internationale vergelijkingen als basis voor concrete stappen om het onderwijs in Nederland te verbeteren en bespreek ik op welke manier gegevens kunnen worden verzameld om in de lacunes te voorzien.

Een belangrijke conclusie is dat we bescheiden moeten zijn met de verwachtingen over internationale vergelijkingen. Op zichzelf is het natuurlijk altijd nuttig om een beeld te hebben over hoe Nederlandse leerlingen het doen in vergelijking met andere landen en de internationale dimensies biedt enkele voordelen boven puur Nederlandse gegevens. Er zijn inmiddels vele internationale indicatoren beschikbaar. Op zichzelf beschouwd is het voor veel van deze indicatoren niet duidelijk hoe de score gerelateerd is aan de kwaliteit of effectiviteit van het onderwijs. De indicatoren moeten daarom in hun onderlinge samenhang worden beschouwd. De grote verschillen tussen de onderwijssystemen maar ook de verschillen in sociaal-economische context tussen landen zijn zo groot dat een eenduidige analyse van wat goed werkt en wat niet, niet eenvoudig is. Daarnaast is leren een proces waarbij activiteiten in de ene fase bepalend zijn voor het leerproces in latere fases en ten slotte voor de einduitkomsten van het onderwijs. Om onderwijs te kunnen verbeteren moet daarom de betekenis van onderdelen voor het geheel begrepen worden. Het soort gegevens dat wordt verzameld voor internationale vergelijkingen levert echter weinig informatie op over dit proces van kennisaccumulatie.

Deze beperkingen worden gereflecteerd in de manier waarop in Nederland gebruik wordt gemaakt van internationale indicatoren. Conclusies op basis van internationale indicatoren die in Nederland naar voren worden geschoven zijn, zijn niet de bevindingen die in het oog springen als alleen naar deze internationale rapporten wordt gekeken. De internationale gegevens worden kennelijk vooral gebruikt ter aanvulling van kwesties die al speelden in de Nederlandse discussie en waarschijnlijk op grond van puur Nederlandse gegevens (of impressies van betrokkenen) aan het licht zijn gekomen.

Ook succesvol onderzoek om de implicaties van bepaalde keuzes bij de inrichting van het onderwijs voor de opbrengsten van het onderwijs te bepalen zijn zowel in Nederland als in andere landen vrijwel altijd gebaseerd op onderzoeksgegevens die betrekking hebben op één land. De betekenis van internationale indicatoren voor wetenschappelijk onderzoek van het onderwijs is dus beperkt.

Hoewel er in Nederland de laatste jaren grote steun lijkt te zijn ontstaan voor evidence based onderwijsbeleid, worstelen de organisaties die op dit front actief zijn nog met de vraag hoe hier vorm aan moet worden gegeven. Een oorzaak hiervoor is een tekort aan goede onderzoeksgegevens. Daarnaast ontbreekt het aan analyses van het kennisaccumulatieproces als geheel op grond waarvan bekeken zou kunnen worden waar de mogelijkheden liggen om door veranderingen van het systeem een substantiële leerwinst te boeken. Besprekingen over mogelijke verbeteringen in het onderwijs worden daarom vaak gevoed door beleidsdiscussies die slechts zijdelings met het onderwijsproces als zodanig te maken hebben, of door toevalligheden in de onderwerpkeuze van wetenschappers.

De econoom Richard Nelson heeft er op gewezen dat de keuze om zaken grondig te onderzoeken een kosten-baten-afweging is. Sommige verbanden zijn heel direct waardoor met weinig data op korte termijn belangrijke verbanden gemeten kunnen worden. Van oudsher is er relatief weinig onderzoek gedaan naar de optimale inrichting van het onderwijs met het oog op het verbeteren van de einddoelen die men hier voor stelt. De reden is dat dit heel lang statisch gezien vrijwel onmogelijk, d.w.z. veel te duur was. De effecten van onderwijs beslaan een periode van vele tientallen jaren en door het grote aantal interveniërende factoren zijn veel observaties nodig om belangrijke verbanden statistisch verantwoord op te kunnen sporen. De kosten-baten-afweging met betrekking tot onderzoek naar de effecten van onderwijs is echter snel aan het verschuiven. Aan de ene kant groeit het belang van onderwijs als input voor de economie en aan de andere kant ontstaan er mogelijkheden voor dataverzameling die goedkoop en uniek zijn in historisch perspectief. Het CBS heeft een zeer goede infrastructuur gecreëerd waarmee een grote diversiteit aan administratieve gegevens en informatie uit steekproefonderzoek aan elkaar gerelateerd kunnen worden, waardoor in de praktijk voor grote aantallen personen gegevens over de onderwijsloopbaan en de latere uitkomsten beschikbaar kunnen komen. Dergelijke gegevens kunnen een enorme stimulans betekenen voor evidence based onderwijsbeleid. Voorwaarden hiervoor zijn echter dat zeer goed op een rij wordt gezet welke aspecten van het onderwijs het meest cruciaal zijn voor de eindopbrengst, zodat het onderzoek zich in eerste instantie kan richten op de factoren die werkelijk de doorslag kunnen geven. Daarnaast moet vanuit een dergelijke kijk op het onderwijsproductieproces als geheel goed nagedacht worden welke gegevens verzameld zullen gaan worden en welke informatie daar op basis van steekproefonderzoek aan toegevoegd dient te worden. Internationale indicatoren kunnen hierbij een functie hebben, maar het valt niet te verwachten dat een substantiële verbetering van de informatie voor het onderwijsbeleid uit de internationale hoek zal komen.

De structuur van de notitie is als volgt. In paragraaf 2 geef ik een overzicht van de internationale vergelijkingen van het onderwijs en bespreek ik de betekenis die dit gehad heeft voor het Nederlandse beleid. In paragraaf 3 ga ik in op de economische betekenis van onderwijs, om vanuit dat perspectief de relevantie van verschillende soorten informatie te kunnen bespreken. In paragraaf 4 wordt ingegaan op de mogelijkheden om einddoelen, inputs en tussendoelen in het onderwijsproces te meten. In paragraaf 5 sluit ik af met enkele conclusies.

2 Internationale vergelijkingen van het onderwijs

De afgelopen tien jaar is er een grote vooruitgang geboekt bij de verzameling van internationaal vergelijkbare gegevens over onderwijs. Het gaat hierbij om uiteenlopende indicatoren.

De meest elementaire informatie betreft gegevens over onderwijsparticipatie. De economen Barro en Lee (1993, 2001) en organisaties zoals de OECD, de ILO en de Europese Commissie/Eurostat hebben belangrijke bijdrages geleverd aan het bij elkaar brengen en vergelijkbaar maken van deze informatie. Het gaat hierbij om gegevens over het aantal jaren gevolgd onderwijs of het bereikte opleidingsniveau.

Een belangrijke vraag hierbij is of het aantal jaren onderwijs of het behaalde opleidingsniveau goede maten voor onderwijsinvesteringen zijn. Zowel binnen een land als tussen landen kunnen er verschillen zijn in wat in een jaar onderwijs geleerd wordt. Kennis ontstaat bovendien niet alleen in formeel onderwijs, maar ook door training en ervaring leren mensen². Om de waarde van jaren onderwijs te kunnen bepalen zijn een aantal onderzoekers het verband gaan schatten tussen beloning en jaren onderwijs en ervaring om zo de waarde van menselijk kapitaal vast te kunnen stellen (bijv. Gollop en Jorgenson, 1980). Blau en Kahn (1996) gebruiken deze methode om de beloningsstructuur van landen met elkaar te vergelijken. Leuven, Oosterbeek en Van Ophem (2004) hebben deze aanpak gerepliceerd, waarbij ze de indicator voor menselijk kapitaal die gebaseerd is op jaren onderwijs en ervaring vervingen door testcores van een test (IALS) die in een groot aantal landen onder de beroepsbevolking was afgenomen. Wat bleek was dat de conclusies volledig veranderden en dus dat de basale gegevens over investeringen in onderwijs duidelijke beperkingen hebben.

Ook bij het gebruik van opleidingsniveau's in plaats van jaren onderwijs speelt de vraag hoe deze tussen landen vergeleken kunnen worden. Methodes die internationale vergelijkingen uitvoeren op basis van het bepaalde opleidingsniveau zijn sterk afhankelijk van de manier waarop opleidingen worden ingedeeld. De standaardclassificatie van opleidingen is de ISCED van UNESCO. De totstandkoming van deze classificatie wordt sterk bepaald door onderhandelingen tussen de verschillende landen. Deze moeten daarbij een afweging maken tussen de behoefte aan objectieve internationaal vergelijkende informatie en de zorg om hun internationale imago. Een

² Borghans, Green en Mayhew (2001) geven een overzicht van de afwegingen die gemaakt moeten worden bij het meten van skills gedurende de gehele loopbaan.

aantal studies (Steedman en McIntosh, 2001, en Freeman en Schettkat, 2001) hebben – door ook een vergelijking te maken tussen de officiële classificatie en de resultaten van de eerder genoemde internationale testen – laten zien dat officiële classificatie niet goed het werkelijke kennispeil van mensen representeert. Met name de lagere ISCED-niveau's blijken in Duitsland (en daarmee waarschijnlijk ook in Nederland) van aanzienlijk hogere kwaliteit te zijn dan dezelfde niveau's in de VS.

Door de analyse van internationale gegevens over het onderwijs werd het duidelijk dat verschillen in onderwijsinvesteringen tussen landen niet te vangen zijn in een enkele indicator. De OECD speelt een leidende rol in het bijeenbrengen van een groot aantal relevante indicatoren van het onderwijs. Deze worden jaarlijks naar buiten gebracht in *Education at a Glance* (zie OECD, 2006). Hierin worden inmiddels zoveel aspecten van het onderwijs bijeengebracht dat een oogopslag al lang niet meer voldoende is om alle gegevens in dit 449 pagina's tellende rapport te overzien.

Al sinds lange tijd bestaat het idee dat voor een goede vergelijking van het onderwijs in verschillende landen, niet alleen naar de inputs in het onderwijsproces, maar ook naar de opbrengsten moet worden gekeken. Uiteraard zijn er gegevens beschikbaar over de arbeidsproductiviteit van werkenden, maar de relatie tussen opbrengsten van het onderwijs en beloning op de arbeidsmarkt wordt mede beïnvloedt door een diversiteit aan andere factoren die niets met het onderwijs te maken hebben. Over andere eindopbrengsten van onderwijs is weinig bekend, ook al omdat er geen consensus lijkt te bestaan om welke opbrengsten het gaat het hoe deze gemeten zouden moeten worden.

Eind jaren '50 namen een aantal onderzoekers het initiatief om voor een goede vergelijking van onderwijs tussen landen direct de kennis van leerlingen te gaan meten. In 1967 werd de IEA (International Association for the Evaluation of Educational Achievement) opgericht. In 1959-1962 voerde de IES een eerste pilot van een vergelijkende studie uit bij 13-jarige leerlingen in 12 landen. Gekeken werd naar de kennis van wiskunde, begrijpend lezen, aardrijkskunde, natuur en techniek, en non-verbale vaardigheden. Op basis van de ervaringen hiermee werd in 1964 de First International Mathematics Study (FIMS) uitgevoerd. IEA organiseerde ook internationaal vergelijkende "science" (natuur en techniek) studies. Op basis van de ervaringen werd de aanpak van deze studies een aantal keer grondig herzien. In de huidige discussie wordt alleen nog gebruik gemaakt van de Third (of Trends in) International Mathematics and Science Study (TIMSS, Mullis et al., 2004 en Martin et al., 2004) en lijkt men soms te vergeten dat er ook al in de jaren '60, '70 en '80 dergelijke vergelijkingen hebben plaatsgevonden.

Naast TIMSS heeft IEA in 2001 ook een vergelijkende studie over lezen uitgevoerd (Mullis et al., 2003). Deze heet PIRLS en is in 2006 herhaald. De OECD en Statistics Canada (2001) hebben het initiatief genomen voor een internationaal vergelijkbare test op het gebied van literacy (geletterdheid) onder volwassenen van alle leeftijden in een groot aantal landen (IALS). Daarnaast organiseert de OECD het zogenaamde PISA-programma, waarbij 15-jarige scholieren op een aantal vakgebieden worden getoetst (OECD, 2001, 2004).

Hoewel de vergelijkbaarheid van de cijfers in de tijd beperkt is, zijn er dus al over een lange periode internationaal vergelijkbare cijfers over onderwijsprestaties beschikbaar. Tabel 1 geeft voor een aantal landen een beeld van de prestaties van 14-jarige leerlingen op het gebied van wiskunde³. Het aantal toetsvragen verschilt sterk tussen de studies waardoor ook de schalen van de indicatoren moeilijk vergelijkbaar zijn. Om vergelijkbaarheid over tijd enigszins mogelijk te maken, zijn alle scores omgezet in verschillen met Nederland ten opzichte van de standaarddeviatie in Nederland. De score van Nederland is daarmee genormaliseerd op 0. Hoewel de omvang van de verschillen varieert over de jaren lijkt er toch een vrij consistent patroon te bestaan. Landen die de laatste jaren beter scoren dan Nederland deden dit meestal ook al in de vroege onderzoeken en visa versa.

Tabel 1

Score van een aantal landen bij internationaal vergelijkbare wiskundetoetsen bij 13- of 14-jarigen (15 jaar bij PISA).

	1964 FIMS	1980 SIMS	1995 TIMSS	1998 TIMSS	2003 TIMSS	2003 PISA
Nederland	0	0	0	0	0	0
Belgie (Vlaanderen)	0.74	-0.08	0.42	0.36	0.02	0.08
Engeland	0.20	-0.39	-0.62	-0.88	-0.76	
Japan	0.89	0.31	1.04	0.78	0.68	0.54
Singapore			1.60	1.28	1.38	
Verenigde Staten	-0.30	-0.32	-0.74	-0.76	-0.64	-1.08
Zweden	-0.50	-0.66	0.22		-0.74	-0.56

Noot: De gerapporteerde score is het verschil van de score met Nederland gedeeld door de standaarddeviatie in Nederland.

Tabel 2

Score van een aantal landen bij internationaal vergelijkbare “science”-toetsen bij 13- of 14-jarigen.

	1970 FISS	1983 SISS	1995 TIMSS	1999 TIMSS	2003 TIMSS
Nederland	0	0	0	0	0
Japan	1.34	0.08	0.24	0.09	0.29
Singapore		-0.65	0.71	0.42	0.76
Verenigde Staten	0.38	-0.65	-0.51	-0.55	-0.16
Zweden	0.39	-0.27	0.22		-0.22

Noot: De gerapporteerde score is het verschil van de score met Nederland gedeeld door de standaarddeviatie in Nederland.

Tabel 2 geeft een vergelijkbaar overzicht voor “science”. In 1970 scoorde Nederland lager dan de drie landen die in dit overzicht zijn opgenomen. In latere jaren lijkt de performance te zijn verbeterd. Bij het maken van deze vergelijkingen valt ook op dat de

³ De gegevens van oudere studies (FIMS en SIMS) zijn gebaseerd op Medrich en Griffith (1992).

groep landen die deelneemt aan een specifiek onderzoek sterk wisselt. Het is daardoor moeilijk om een constante vergelijkingsgroep te nemen en een aantal landen dat als vergelijking voor de hand had gelegen kon om die reden niet worden opgenomen.

Bij TIMSS worden leerlingen veelal in verschillende leeftijdsgroepen getoetst. Door de scores tussen deze groepen te vergelijken ontstaat een beeld over de relatieve groei van het niveau van leerlingen in de tussenliggende jaren. Tabel 3 laat zien dat België en de Verenigde Staten achterblijven ten opzichte van Nederland bij de vergelijking van 10- en 14-jarigen, maar dat kinderen in Singapore en Japan een sterke ontwikkeling in science doormaken.

Tabel 3

Score van een aantal landen bij TIMSS 2003 op “science” bij 10- (grade 4) en 14-jarigen (grade 8).

	grade 4	grade 8	verschil
Singapore	594	605	11
Japan	565	570	5
Belgie (Vlaanderen)	551	537	-14
Nederland	540	536	-4
Verenigde Staten	518	504	-14

In het algemeen haalt Nederland hoge gemiddelde scores in de internationaal vergelijkende testen. Bij PISA 2003 scoort Nederland hoog op het gebied van de wiskunde. Alleen Hong-Kong, Finland en Korea scoren hoger. Bij lezen staan slechts 8 van de 40 landen hoger dan Nederland. Bij science zijn dat er 7. Bij de laatste TIMSS zijn bij wiskunde alleen de scores van Singapore, Zuid Korea (alleen bij grade 8), Hong Kong, Taiwan, Japan en Vlaanderen beter dan die van Nederland. Bij natuur en techniek halen bij grade 8 ook Estland, Hongarije hogere resultaten, maar zijn die van Vlaanderen lager dan Nederland. Nederland scoort iets minder hoog bij natuur en techniek in grade 4. Daar staan ook Engeland, Litouwen en Rusland boven Nederland. In PIRLS heeft alleen Zweden een gemiddeld hogere score dan Nederland en bij IALS staan alleen Zweden, Noorwegen, Denemarken en Finland hoger in de ranking. Wat opvalt is dat in een aantal testen (wiskunde en natuur en techniek) een aantal Aziatische landen de top blijkt te vormen, terwijl bij taal Scandinavië de boventoon voert. Nederland behaalt over de volle breedte zeer goede resultaten.

In Education at a Glance brengt de OECD (2006) naast deze internationaal vergelijkende testcores een groot aantal andere indicatoren over het onderwijs bijeen om de performance in de verschillende landen te kunnen vergelijken. Als dit rapport wordt doorgenomen om te kijken wat de positie van het Nederlandse onderwijs in internationaal perspectief is, vanuit de nadrukken die de OECD legt, komt het volgende beeld naar voren:

- Nederland scoort hoog bij de wiskundeprestaties van leerlingen. De verschillen in de prestaties zijn voor een groot deel verschillen tussen scholen en niet binnen scholen. Het percentage leerlingen dat heel laag scoort is opmerkelijk laag.

- De beoogde instructietijd van het Nederlandse onderwijs is opvallend hoog in internationaal perspectief. Als de beoogde instructietijd een goede reflectie is van de werkelijke instructietijd dan krijgen Nederlandse leerlingen dus veel meer les dan leerlingen in andere landen. Ook het deel van de instructietijd dat besteed wordt aan taal en rekenen is met name op de basisschool erg hoog. “Science” daarentegen krijgt in Nederland op de basisschool opmerkelijk weinig aandacht.
- Bij een groot aantal indicatoren wijkt Nederland niet sterk af van het gemiddelde. Dat betreft onder andere de arbeidsmarktparticipatie van lager opgeleiden; de lonen van leraren in PPP (als percentage van het GDP zijn de lonen van de leraren aan de lage kant); het aandeel van de private bijdragen aan het onderwijs; de niet-formele training van volwassenen; klassengrootte en gebruik van ICT.
- Nederland weet dus lage kosten met goede resultaten te combineren volgens de OECD: “Lower unit expenditure does not necessarily lead to lower achievement and it would be misleading to equate lower unit expenditure generally with lower quality of educational services. For example, the cumulative expenditure per student between primary and secondary education of Korea and the Netherlands are below the OECD average and yet both were among the best-performing countries in the PISA 2003 survey.”
- De OECD is niet heel tevreden met de steun die leerlingen in de klas krijgen: “In Hong Kong-China, high expectations for students go hand in hand with a highly supportive learning environment where students consistently report that teachers show an interest in every student’s learning, give extra help when students need it, and continue teaching until all students understand. In contrast, students in all of Europe’s major economies, and most notably Germany, France, Italy, Poland and the Netherlands, reported significantly lower levels of teacher support. The lack of ambition combined with a lack of support in Europe’s education system is a troublesome base for the future success of its people.” (p. 18).
- Opvallend laag scoort Nederland met betrekking tot het gemiddeld aantal jaren dat mensen in het onderwijs zitten. Vooral dit maakt het Nederlandse onderwijs relatief goedkoop.

Ook blijkt dat Nederland vaak ontbreekt in de overzichten. Op grond van de beoordeling van de OECD is het relatief vaak niet goed mogelijk gegevens over Nederland op een betrouwbare manier vergelijkbaar te maken met de gegevens van de andere landen.

Deze internationale gegevens zijn ook door Nederlandse instituten gebruikt om de positie van Nederland op onderwijsgebied in kaart te brengen. Het Centraal Planbureau (CPB) heeft onlangs een rapport opgesteld waarin het Nederlandse onderwijs (en onderzoek) wordt belicht vanuit deze internationale indicatoren (Antenbrink et al., 2005). Bekeken wordt wat de kwaliteit van het Nederlandse onderwijs is en in hoeverre internationale indicatoren richting kunnen geven aan het beleid. Het CPB komt ook tot de conclusie dat het Nederlandse kennissysteem niet beter of slechter functioneert dan dat van de kopgroep in Europa. In grote lijnen volgen vergelijkbare analyses als hierboven zijn opgesomd. Hierbij gaat men uitgebreid in op het feit dat in Nederland op relatief jonge

leeftijd de arbeidsmarkt wordt betreden. Vooral het percentage jongeren dat met een kwalificatie lager dan ISCED 3 de arbeidsmarkt betreedt baart het CPB zorgen⁴.

Interessant is dat de OECD helemaal niet tot deze conclusie komt. Zij schrijft zelfs (p. 16): “It is also noteworthy that most of East Asia’s education systems succeed without leaving many students behind, even if Finland, Canada and the Netherlands also did well in this respect.” Deze conclusie van de OECD lijkt vooral gebaseerd op het lage percentage leerlingen met een lage wiskundescore bij PISA, terwijl het CPB kijkt naar de diploma’s van degenen die het onderwijs verlaten. Zonder een oordeel te willen vellen over welke conclusie terecht is, is het interessant om te constateren dat dezelfde set indicatoren ruimte laat om Nederland op het gebied van de minst succesvollen in het onderwijs zowel negatief als positief te beoordelen.

Een andere opvallende conclusie is dat in de ogen van het CPB internationale indicatoren niet erg nuttig zijn voor het beleid. “internationale benchmarks zijn waardevol in de beleidsvoorbereiding”. “Een vergelijking van indicatoren is een goed startpunt voor verdere analyse”, maar “levert geen concrete beleidsrecepten.” In het rapport worden een aantal problemen besproken waar men mee te maken heeft bij het gebruik van de internationale indicatoren voor het beleid. Experimenten en beleidsevaluaties binnen Nederland zijn volgens het CPB van veel groter belang dan vergelijkingen tussen landen, omdat steeds onduidelijk blijft in hoeverre de omstandigheden in andere landen vergelijkbaar zijn met de Nederlandse situatie.

In Cornet et al. (2006) formuleert het CPB kansrijke maatregelen voor de verbetering van het Nederlandse onderwijs. Hierbij probeert men het principe dat onderwijsbeleid “evidence based” moet zijn inhoud te geven. Het streven is hierbij om de beoordelingen te baseren op harde evidence. Dat wil zeggen op onderzoeksresultaten die gebaseerd zijn op experimenteel of quasi-experimenteel onderzoek (dat is onderzoek waarbij gebruik wordt gemaakt van een treatment-groep en een control-groep die door toevalligheden zijn ontstaan). Hoewel het van groot belang is dat deze vertaalslag wordt gemaakt, maakt dit rapport duidelijk dat het niet eenvoudig is om concreet inhoud te geven aan het idee om onderwijsveranderingen te baseren op harde onderzoeksresultaten. Het basisprobleem van dit rapport is met name het ontbreken van een goed startpunt van de analyse. Er zijn heel veel facetten van het onderwijsproces die in principe aangepast kunnen worden en er moet dus een keuze gemaakt worden welke facetten prioriteit moeten krijgen. De keuze van het CPB wordt in hoge mate gestuurd door beleidsmogelijkheden waar onderzoeksresultaten voor beschikbaar zijn of die veel genoemd zijn in de maatschappelijke discussie. Er is echter weinig reden om aan te nemen dat de keuzes die wetenschappers maken bij hun analyses een goede reflectie vormen van de opties met de grootste kansen voor het beleid. Hun keuzes worden gestuurd door het streven om in toptijdschriften te publiceren. Beschikbaarheid van kwalitatief goede data en een ingenieus onderzoeksdesign zijn daarbij belangrijker dan de ingeschatte efficiëntiewinsten in het onderwijs. De thema’s die uit de maatschappelijke discussie voortkomen betreffen vooral beleidsdiscussies waarbij een belangentegenstelling voorop

⁴ Ook de samenvatting van Education at a Glance 2006 die de Minister naar de kamer heeft gestuurd benadrukt sterk dit aspect.

staat. Het gaat daardoor meer over de vraag wie voor het onderwijs moet betalen dan om de vraag of de onderwijsproductiviteit als zodanig verbeterd kan worden.

Om houvast bij deze afwegingen te krijgen baseert het CPB zich vooral op de economische redenering van “externaliteiten” en “marktfalen”. Zoals ik in de volgende paragraaf zal bespreken gaat deze aanpak er van uit dat betrokkenen in het onderwijs in principe optimale beslissingen kunnen nemen. Scholen en leraren weten wat goed is voor leerlingen; Leerlingen weten wat goed voor hen is. In een dergelijke situatie hoeft de overheid er alleen voor te zorgen dat degene die betaalt ook degene is die de baten krijgt. De grote verschillen in de manier waarop landen invulling geven aan het onderwijs maakt echter duidelijk dat er nog maar weinig bekend is over wat in het perspectief van een gehele levensloop een efficiënte inrichting van het onderwijs is. We moeten er dus rekening mee houden dat mensen (en scholen) niet weet wat de beste keuzes zijn. Experimenteel onderzoek is juist nodig om deze kennis te verwerven. Redeneringen die gebaseerd zijn op marktwerkingsprincipes om de keuzes voor experimenteel onderzoek voor het verbeteren van het onderwijs te funderen leidt daarom tot een conceptuele spanning. Een grondige kennis van hoe het leerproces in Nederland verloopt om zo de cruciale parameters te achterhalen die bepalend zijn voor het succes van het systeem is nodig om verantwoorde keuzes te maken voor evidence based onderwijsbeleid.

Naast het CPB heeft ook de onderwijsraad (2005) de positie van het Nederlandse onderwijs in internationaal perspectief geschetst om te achterhalen voor welke uitdagingen het beleid staat. De onderwijsraad heeft de vergelijking tussen Nederland en andere landen teruggebracht tot een vergelijking van Nederland en een referentieland, waarbij het moeilijk te doorgronden is waarop de informatie over dit referentieland is gebaseerd. In grote lijnen stemmen de bevindingen in dit rapport overeen met die van het CPB: De verschillen tussen Nederland en dit referentieland zijn niet zo groot, alleen zijn Nederlandse leerlingen opvallend goed in wiskunde en lezen en halen relatief weinig leerlingen het ISCED 3 niveau. In tegenstelling tot het CPB-rapport wordt niets gezegd over de mogelijkheden om deze cijfers te vertalen naar concrete beleidsdoelen en bij de adviezen voor het beleid die worden gegeven, is het verband met de internationale vergelijkende cijfers zeer onduidelijk. Discussies die binnen Nederland zijn ontstaan over mogelijkheden het onderwijs te verbeteren lijken de leidraad te vormen voor deze adviezen, en de internationale cijfers worden gebruikt om deze discussie te illustreren. Dit wijst er wederom op dat – hoewel het in theorie belangrijk lijkt om een internationaal perspectief te gebruiken om de positie van Nederland te kunnen schetsen – het in de praktijk zeer moeilijk is de veelheid aan internationale indicatoren te interpreteren, zonder daarbij de Nederlandse discussie als vertrekpunt te nemen. Gegevens in lijn met de oordelen die er over Nederland bestaan worden daarmee bevestigd, terwijl indicatoren die hiermee in strijd lijken te zijn als moeilijk interpreteerbaar aan de kant worden geschoven.

Het Sociaal-Cultureel Planbureau (2007) probeert een beeld te geven van het positie van het Nederlandse onderwijs door de veelheid aan gegevens uit internationale bronnen terug te brengen tot enkele indicatoren. Op grond daarvan concludeert men dat de educatieve vaardigheden die in Nederland worden geproduceerd in redelijke verhouding

staan tot de kosten die ervoor worden gemaakt (Nederland zit dicht bij de “production frontier”) en dat het vertrouwen dat Nederlanders hebben in het onderwijs in lijn is met het geleverde product. Om tot deze conclusie te komen moet het SCP de veelheid aan indicatoren terugbrengen tot zulke abstracte vormen dat concrete conclusies over beleidsaanbevelingen niet meer mogelijk lijken te zijn.

Het blijkt dus praktisch een probleem te zijn om de informatie op grond van internationale indicatoren om te zetten in concrete aanbevelingen voor veranderingen in het onderwijs. Bij de bestudering van de grote hoeveelheid indicatoren ontstaat er geen eenduidigheid over landen die het goed doen en landen die het slecht doen. Als er al duidelijke verschillen in patronen bestaan dan betreft dat het verschil tussen rijkere en armere landen en het verschil tussen Westerse en Aziatische landen. Terwijl de basisgedachte van de internationale vergelijking is dat landen die het minder goed doen kunnen leren van de landen die het beter doen, ontstaat er bij de praktische toepassing van dit principe twijfel of de landen die zo hoog scoren wel echt vergelijkbaar zijn met het eigen land. Verschillen tussen landen die meer vergelijkbaar lijken te zijn, zijn echter vaak klein en minder eenduidig. Het ene jaar scoort het ene land hoog en het andere laag en op de ene indicator doet het ene land het goed terwijl een ander land op een andere indicator beter scoort.

3 De economische betekenis van onderwijs

Om op een zinvolle manier na te denken over de vraag op welke wijze internationale vergelijkingen bijdragen bij keuze over veranderingen in het onderwijs, moet bekeken worden wat de rol van onderwijs is en op welke wijze verschillen tussen individuen en landen tot stand komen. Cruciaal hierbij is dat het gaat om een economisch proces. Mensen leren niet puur om zoveel mogelijk te weten te komen, maar volgen onderwijs om daarmee kennis en vaardigheden te verwerven die hen naderhand voordelen opleveren. Bij deze voordelen wordt meestal vooral gekeken naar betere mogelijkheden op de arbeidsmarkt, maar het kan ook gaan om bredere doelen zoals persoonlijke ontplooiing, een gezonder leven, etc. Ook bij het werk kan zowel gekeken worden naar de beloning, als naar de tevredenheid met het werk. Voor een deel kan het gaan om baten die toevallen aan het individu, maar er kunnen ook baten van onderwijs zijn waar het individu zelf geen direct voordeel van heeft, maar waar anderen voordeel van hebben. Er wordt daarom een onderscheid gemaakt tussen individuele baten van onderwijs en de maatschappelijke baten (dat zijn zowel de individuele baten als baten voor anderen).

Voor de redenering over de betekenis van onderwijs en de manier waarop het beleid aan het onderwijs vorm kan geven is het niet van belang wat de doelen van het onderwijs zijn. Naarmate het onderwijsbeleid zich meer gaat richten op de inhoud van het onderwijs wordt het natuurlijk wel cruciaal om een concrete invulling aan deze doelen te geven.

Omdat onderwijs van belang is om bepaalde doelen te realiseren beschouwen economen onderwijs als een investering. Door tijd en middelen in onderwijs te steken nemen de kennis en vaardigheden toe, waarvan men naderhand kan profiteren. Een groot verschil

tussen deze investeringen in menselijk kapitaal en investeringen in fysiek kapitaal is dat tijd de belangrijkste input in het leerproces is. Goede onderwijsmiddelen en goede leerkrachten zijn uiteraard van belang maar een cruciale input in het leerproces is altijd de tijd die de leerling aan leren besteed. Door voldoende mensen in te huren kan men in principe in vrij korte tijd een grote fabriek laten bouwen, maar om een grote hoeveelheid menselijk kapitaal te ontwikkelen moet de betrokken persoon zelf voldoende tijd hiervoor beschikbaar maken.

Omdat onderwijs een investering is, geldt in principe hoe eerder hoe beter. Immers als men eerder de investeringsperiode af kan ronden, kan ook eerder geprofiteerd worden van de baten van het onderwijs en is de terugverdienperiode langer. Dat geldt ook voor de opbrengsten buiten de arbeidsmarkt. Als mensen eerder leren om gezond te leven kunnen ze ook hier langer van profiteren.

Wat leerlingen leren op school is uiteraard geen homogeen goed. Er zijn verschillende vormen van kennis en bepaalde kennis van weer bijdragen in het verkrijgen van andere vormen van kennis. Men kan daarom basiskennis definiëren als de kennis die een grote rol speelt bij het latere leerproces. Leren lezen en schrijven is hier natuurlijk een duidelijk voorbeeld van. Als een kind eenmaal kan lezen en schrijven kan het zelf zelfstandig meer lesstof tot zich nemen en kunnen andere vakken sneller geleerd worden. Voor een optimaal eindrendement van het onderwijs kan de volgorde waarin dingen geleerd worden dus cruciaal zijn.

Daarnaast is echter ook de natuurlijke ontwikkeling van een kind van invloed op het vermogen om te leren. Het is bekend uit de psychologische literatuur dat bepaalde vaardigheden makkelijker geleerd worden in bepaalde periodes van het leven. Heel vroeg beginnen met lezen en schrijven heeft een relatief laag rendement. Pas vanaf een bepaalde leeftijd kunnen kinderen vrij makkelijk lezen en schrijven onder de knie krijgen. Jonge kinderen leren daarentegen makkelijker dingen uit hun hoofd en hebben een goed ontwikkeld vermogen nieuwe talen te leren en krijgen de uitspraak daarvan beter onder de knie. Cunha en Heckman (2006) beschouwen het leerproces als een productieproces en spreken van sensitive en cruciale periodes voor het leerproces.

De opbrengst van het leerproces hangt ook niet alleen af van de opbouw van het curriculum, maar wordt ook bepaald door de non-cognitieve kenmerken van de leerling, zoals doorzettingsvermogen en zelfvertrouwen. Borghans et al. (2007) geven een overzicht van de economische en psychologische literatuur over deze invloed van non-cognitieve vaardigheden en laten zien dat de ontwikkeling van persoonskenmerken zelf ook een investeringsproces is.

In theorie geldt dus dat als de doelen van onderwijs duidelijk zijn er een optimaal curriculum vastgesteld kan worden – dat voorschrijft in welke volgorde welke lesstof aan de orde zou moeten komen en rekening houdt met de non-cognitieve ontwikkeling – waarbij de gewenste eindopbrengst wordt gemaximaliseerd.

Traditioneel gezien zijn economen bij hun analyse van het onderwijs niet geïnteresseerd in de invulling die het lesprogramma krijgt. Ze kijken bij het effect van onderwijs alleen naar het aantal jaren dat mensen op school zitten en eventueel de financiële input (kosten van leraren, gebouwen en materialen). De reden hiervoor is niet dat men denkt dat ieder jaar onderwijs hetzelfde is en het er niet toe doet wat mensen op school doen. Het theoretische argument dat wordt gehanteerd voor een strikte scheiding tussen de beslissing hoeveel onderwijs wordt genoten en de beslissingen over de inhoud van het programma zijn gebaseerd op de veronderstelling dat leraren gegeven de middelen en lesjaren die er beschikbaar zijn het optimale lesprogramma zullen selecteren (zie Borghans en Heijke, 2005).

De overheid hoeft in dat geval alleen te beslissen hoeveel men investeert in onderwijs. Omdat de individuele baten van onderwijs kleiner kunnen zijn dan de maatschappelijke baten, zullen individuen onderinvesteren in onderwijs en is steun van de overheid nodig om het optimum te bereiken. Als dat eenmaal besloten is zal binnen het onderwijs – zo is de impliciete veronderstelling – optimaal gebruik worden gemaakt van de middelen om er zoveel mogelijk uit te halen. Deze manier van kijken is de ultieme versie van beleid dat pleit voor een overheid op afstand. In het beleid vindt men deze gedachtegang terug in de stelling dat de leraar (of de school) het beste weet wat goed is voor de leerlingen.

Als een land bij een internationale vergelijking zoals PISA of IALS relatief slecht scoort betekent dat vanuit deze optiek in de eerste plaats dat men kennelijk te weinig investeerde in onderwijs. De Duitse reactie op de lage PISA-scores om meer in het onderwijs te investeren past in dit denkkader.

De vraag is natuurlijk of het inderdaad vanzelfsprekend is dat leraren/scholen de beschikbare middelen optimaal benutten. Een eerste reden om hieraan te twijfelen is dat de incentives die leraren en scholen hebben mogelijk niet parallel lopen met de langetermijn-belangen van leerlingen. Uit onderzoek komt naar voren dat dit inderdaad niet het geval is en er zijn zelfs aanwijzingen dat de keuzes die ouders voor hun kinderen maken niet altijd ook het belang van de kinderen zo goed mogelijk weerspiegelen (Jacob en Lefgren, 2005).

De beleidsinstrumenten die hier uit voortkomen richten zich op het verbeteren van de incentives en competitie. Hier is veel over geschreven en hoewel moeilijk analyseerbaar redelijk veel onderzoek naar gedaan. Hanushek (2003) en Lavy (2004) laten zien dat betere incentives voor leraren leidt tot betere onderwijsprestaties. Hoxby (2000) en Lavy (2006) laten zien dat meer competitie tussen scholen de prestaties van leerlingen ten goede komt. Er worden echter ook resultaten gevonden die het belang van competitie en duidelijkere incentives niet lijken te ondersteunen (Rothstein, 2005, zie Hoxby 2005 voor een reactie). Jacob en Levitt (2003) laat zien dat high stakes testing in Chicago heeft geleid tot meer fraude door leraren bij de eindexamen om hun leerlingen door de toets heen te slepen. Urquiola en Hsieh (2002) laten zien dat beloningen voor scholen op basis van eindexamenresultaten leiden tot teaching to the test.

Ook internationale data zijn gebruikt om de rol van incentives te onderzoeken. Jorges en Schneiders (2004) vinden de effecten van concurrentie tussen scholen op basis van gegevens van TIMSS. Wossmann (2005) laat zien dat onderwijssystemen met een centraal eindexamen betere resultaten laten zien dan andere onderwijssystemen. Het idee hierbij is dat een centraal examen de resultaten van scholen transparanter maakt en daarmee de concurrentie bevordert. Het grote voordeel van de internationaal vergelijkbare testen zoals PISA is dat deze losstaan van de specifieke toetsen in de afzonderlijke landen en er bij het maken van de toetsen geen grote belangen zijn (of lijken te zijn) voor de scholen en leerlingen. Schijnverbeteringen in het onderwijs doordat scholen zich in hun lesprogramma meer gaan richten op de centrale examens zullen daardoor door deze testresultaten niet beïnvloeden. Vooral als internationale testen met een vaste frequentie zouden worden herhaald, dan geeft dat de mogelijkheid om de trends in landelijke scores te vergelijken met de trends in deze internationale metingen.

Het licht voor de hand dat niet alleen problemen rondom incentives maar ook onzekerheid over op welke manier het onderwijs het beste ingericht kan worden om optimale opbrengsten te genereren een reële zorg is voor het onderwijsbeleid. De implicaties van keuzes in het onderwijs hebben vrijwel altijd betrekking op de lange termijn, waarbij een groot aantal invloeden door elkaar lopen. Oorzaak en gevolg zijn daarom empirisch moeilijk vast te stellen. De manier waarop onderwijs in verschillende landen vorm heeft gekregen is dan ook voor het overgrote deel gebaseerd op veronderstellingen over de betekenis van onderdelen voor het geheel. Als er al sprake is van optimalisatie op basis van empirische toetsing dan betreft dit korte-termijn-effecten en gaat het over de vraag hoe binnen een vak of cursus bepaalde leerdoelen zo goed mogelijk bereikt kunnen worden. Murnane en Nelsen (2005) in dit opzicht een interessant onderscheid tussen de basale structuur van het onderwijsproces en de specifieke aanpak van een school en manier van lesgeven van een leraar. Met betrekking tot de basale structuur kan niet verwacht worden dat scholen of leraren weten welke manier van lesgeven optimaal is in het licht van de totale schoolloopbaan. Hier zijn gerandomiseerde experimenten daarom het ideaalbeeld voor onderzoek. De specifieke aanpak van het onderwijs is echter moeilijk te vangen in algemene instructies. Hier komt het aan op het talent van de leraar of de kwaliteit van de school. In de ogen van Murnane en Nelson zijn dit de aspecten die van een gewone school een “high-performance” school kunnen maken.

Weinig is bekend of de effecten van de inrichting van het onderwijs voor de ontwikkeling van een leerling op langere termijn en weinig is bekend over de implicaties die verschuivingen in het belang van bepaalde kennis en vaardigheden of andere onderwijsopbrengsten voor de inrichting van het curriculum. De keuzes die landen hebben gemaakt zijn gebaseerd op gezond verstand en verschillen sterk van land tot land. De oprichters van IEA hadden destijds dan ook het beeld voor ogen dat deze verschillen in aanpak van het onderwijs tussen landen gezien kan worden als een natuurlijk laboratorium. Door de opbrengst van het onderwijs in de verschillende landen met elkaar te vergelijken, zou een beeld ontstaan van succesvolle en minder succesvolle benaderingen en zouden landen van elkaar kunnen leren.

Belangrijke vragen waar we nauwelijks een antwoord op weten zijn: Op welke leeftijd zouden kinderen moeten beginnen met leren lezen en schrijven? Wat is het belang van voorschoolse educatie voor de ontwikkeling van kinderen? Hoeveel tijd zou er gedurende het gehele curriculum aan rekenen en wiskunde moeten worden besteed? En is het belangrijk dat vroeg in de schoolloopbaan de nadruk op deze vakken wordt gelegd? Is het van belang dat gymnasiasten nog steeds Grieks en Latijn leren? Is het belangrijk om computervaardigheden te leren op school? Wat zijn de effecten van niveaudifferentiatie in het middelbaar onderwijs?

Bij al deze vragen gaat het om effecten van keuzes in het onderwijs op het functioneren van mensen op de langere termijn. Het gaat dus niet om de vraag op welke manier je het beste leert lezen, rekenen, enz. De werkelijke betekenis van onderwijs zit in de uiteindelijk consequenties voor de ontwikkeling van mensen over hun hele leven. Zo is bekend dat het vermogen van kinderen om te leren lezen en schrijven sterk afhankelijk is van leeftijd. Gemiddeld gesproken is een zesjarig kind hier meer aan toe dan een vijfjarig kind. De vraag die echter moeilijker te beantwoorden is, is hoeveel het waard is om ondanks dit verschil toch iets vroeger te beginnen met leren lezen en schrijven. Het leren zal langer duren en de opbrengst zal wellicht lager zijn, maar leerlingen zullen ook eerder deze vaardigheden bezitten en dus langer kunnen profiteren van de verworven kennis. De afweging die optimaal is vanuit economisch perspectief is daardoor niet gelijk aan de afweging die optimaal is vanuit een psychologisch of onderwijskundig perspectief. Psychologisch onderzoek over het functioneren van de hersenen bij leeractiviteiten, moet dus gecombineerd worden met economisch onderzoek over de waarde van wat geleerd wordt in specifieke fases van de schoolloopbaan.

Naast het belang van timing speelt hierbij ook dat meer aandacht besteden aan één vak doorgaans betekent dat er minder aandacht wordt besteed aan een ander vak⁵. Daarnaast kunnen activiteiten in het onderwijs ook een substituut zijn voor leerprocessen buiten het onderwijs. Kinderen leren ook thuis van hun ouders. Meer tijd op school betekent dus minder leertijd in de thuisomgeving. Daarnaast leren kinderen ook in hun vrije tijd. Vooral bij sociaal-emotionele ontwikkeling is daardoor steeds de vraag of aandacht op school een vervanging is voor ontwikkeling die toch al zou hebben plaatsgevonden. Cunha en Heckman (2007) laten zien dat de rol van ouders in het leerproces inderdaad zeer groot is en de school zeker niet alleen bepalend is voor de uitkomsten van de ontwikkeling van kinderen. Ten slotte leren mensen ook heel veel door te werken. Met name de eerste jaren op het werk zijn zeer ervaringsrijk (Borghans, Golsteyn en De Grip, 2006). Door op school veel aandacht te besteden aan zaken die men zich op het werk vrij makkelijk eigen zou maken, wordt de overgang tussen school en werk verbeterd, maar is er feitelijk juist sprake van een inefficiëntie.

Welke functie hebben internationaal vergelijkbare indicatoren bij het verbeteren van het onderwijs vanuit dit perspectief? De rol van een losstaande indicator is beperkt. Als wordt gemeten hoe goed leerlingen in verschillende landen op 15-jarige leeftijd zijn in

⁵ Bij een verschuiving in accenten in het leerproces hoeft de verhouding tussen meer aandacht voor het ene vak en minder aandacht voor een ander vak niet 1 op 1 te zijn, doordat geïntegreerde lesmethoden meer op kunnen leveren dan de som van twee afzonderlijke lessen. Er zal echter altijd sprake zijn van enig verlies.

wiskunde dan kunnen verschillen in scores het gevolg zijn van een andere opbouw van het curriculum. Wellicht begint men in het ene land pas later met wiskunde en is daardoor het niveau van leerlingen lager dan in een ander land. De Nederlandse situatie lijkt goed in dit beeld te passen. In Nederland wordt veel tijd op de basisschool besteed aan rekenen en taal daardoor zou de consequentie kunnen zijn dat scholieren op 15-jarige leeftijd een voorsprong hebben op leerlingen in andere landen.

De cruciale vraag of het ook nuttig is om zo vroeg in de schoolloopbaan de nadruk op rekenen, wiskunde en taal te leggen wordt daarmee niet beantwoord. Een vergelijkbare situatie doet zich voor bij de hoge drop-out rates van Nederlandse scholieren. Wordt het probleem van de jongeren die zonder startkwalificaties de arbeidsmarkt betreden overschat omdat uit PISA blijkt dat zelfs deze jongeren het internationaal gezien nog goed doen, of is sterke nadruk die vroeg in het Nederlandse systeem ligt op wiskunde en taal wellicht de reden dat deze jongeren zijn afgehaakt?

Ook is het onduidelijk hoe belangrijk het is om een bepaald wiskundeniveau te behalen en in hoeverre dit hoge wiskundeniveau ten koste is gegaan van andere (wellicht niet gemeten) prestaties van leerlingen. In onderzoek zou geprobeerd kunnen worden om de scores te corrigeren voor verschillen in omstandigheden (zoals opleidingsniveau ouders en GDP per capita van het land) en de opbouw van het curriculum. Wat resulteert, is dan een indicator over hoe goed het onderwijs presteert in het genereren van wiskundekennis bij gelijke inzet van middelen. De vraag waarom het ene land slechter presteert dan het andere is op grond van deze cijfers moeilijk te beantwoorden. Deze aanpak geeft ook geen antwoord op de vraag wanneer leerlingen het beste wiskunde kunnen leren en hoeveel wiskundekennis optimaal zou zijn. Om deze vragen te analyseren is een andere invalshoek vereist. Hiervoor zou een analyse gemaakt kunnen worden van de samenhang tussen wiskundeniveau van 15-jarigen en hun later functioneren op de arbeidsmarkt. Bij de eerste invalshoek worden de verschillen in prestaties in de niveaus van leerlingen toegeschreven aan de kwaliteit van het onderwijs. In het tweede geval worden ze gezien als een indicator voor keuzes die in het onderwijs gemaakt zijn.

Een aantal onderzoekers hebben vanuit deze invalshoek internationaal vergelijkbare gegevens benut om de effecten van een bepaalde opzet van het curriculum te analyseren. Hanushek en Wossman (2006) vinden negatieve effecten van tracking. Hanushek en Luque (2003) onderzoeken het verband tussen inputs en efficiencyproblemen van het onderwijs. Ook de OECD maakt in haar rapporten vaak vergelijkingen die op dit principe zijn gebaseerd en laat bijvoorbeeld zien dat bij systemen waar zittenblijven veel voorkomt de band tussen sociale herkomst van de ouders en prestaties van de leerling groter is.

Vanuit een onderzoeksperspectief zijn deze analyses echter niet ideaal. De analyse komt feitelijk neer op de vergelijking van een tien- of twintigtal punten (één per land) waarbij het verband vaak verre van eenduidig is. Door het geringe aantal observaties is de statistische betrouwbaarheid van deze analyses laag. Bovendien kan in de praktijk maar naar een of eventueel twee aspecten van het onderwijsbeleid simultaan worden gekeken. De veronderstelling die daarom gemaakt moet worden is dat deze aspecten niet

samenhangen met andere aspecten van het curriculum of land-specifieke omstandigheden die niet in de analyse worden betrokken.

In de praktijk van het onderzoek blijken meer robuuste analyses om de effecten van een bepaalde invulling van het curriculum te analyseren altijd gebaseerd te zijn op gegevens over één land. Het kan hierbij zowel gaan om variatie tussen leerlingen binnen een systeem of om veranderingen in het systeem als gevolg van veranderingen in het beleid. Er zijn twee voordelen. Door naar individuen in plaats van naar landen te kijken komen er veel meer observaties beschikbaar, waardoor de statistische betrouwbaarheid sterk toeneemt. Belangrijker is echter nog dat de verschillen die zich binnen een land voordoen, spelen tegen de achtergrond van voor de rest gelijke omstandigheden. Bij de analyse van de verschillen tussen onderwijsaspecten spelen interfererende systeemverschillen dus geen rol.

Het is overigens opmerkelijk dat ook de internationaal vergelijkende onderzoeken PISA, TIMSS, PIRLS en IALS vaak gebruikt worden voor afzonderlijke analyses per land. Ammermuller, Heijke en Wossmann (2005) bestuderen stimulerende factoren voor het onderwijs in Oost-Europese landen. Ammermuller en Pischke (2006) onderzoeken peer-effecten in het basisonderwijs en Wossmann en West (2006) gebruiken TIMSS voor de bestudering van het effect van klassengrootte op schoolprestaties. Dit reflecteert dat analyses van verschillen binnen een land vaak beter geschikt zijn om onderzoeksvragen te beantwoorden. Daarnaast zijn deze internationale databestanden voor onderzoekers vrij beschikbaar, terwijl het in veel gevallen zeer moeilijk kan zijn om toegang te krijgen tot data van een specifiek land. Ook in Nederland zijn veel relevante onderzoeksbestanden niet erg toegankelijk voor onderzoekers⁶.

Van groot belang bij dergelijke analyses is dat er sprake is van exogene variatie. Dat wil zeggen dat de verschillen die tussen individuen worden waargenomen niet het gevolg moeten zijn van keuzes, maar dat ze gerelateerd moeten zijn aan verschillen in omstandigheden die buiten de invloed van het individu liggen en ook niet samenhangen met andere factoren die een rol spelen in het onderwijsproces.

Een verandering in beleid is een voorbeeld. Als leerlingen niet kunnen kiezen of ze gebruik maken van de oude of de nieuwe regeling, de omstandigheden buiten deze beleidswijziging nagenoeg gelijk blijven, en de verdeling van karakteristieken van leerlingen voor en na de verandering nagenoeg gelijk zal zijn, geeft een vergelijking van voor en na de beleidsverandering een zuiver beeld van de effecten. Probleem hierbij kan nog zijn dat het effect van de beleidsveranderingen niet helemaal te onderscheiden is van de geleidelijke ontwikkelingen die autonoom plaatsvinden, dat beleidsmaatregelen vaak een veelheid aan veranderingen tegelijk bewerkstelligen en dat leerlingen of instellingen in hun gedrag vooruit kunnen lopen op veranderingen of juist activiteiten uit kunnen stellen om van de nieuwe regeling gebruik te kunnen maken.

Voor het vaststellen van effecten van de inrichting van het onderwijs is een gerandomiseerd experiment ideaal. Het is in Nederland tot nog toe echter zeer

⁶ Overigens is deze situatie bij de overgang van WSA naar DANS voor een aantal bestanden verbeterd.

ongebruikelijk om weloverwogen de inrichting van het curriculum te variëren, om zo effecten hiervan te kunnen identificeren. Zelfs bij initiatieven waar dit heel makkelijk zou kunnen omdat er wegens gebrek aan middelen minder scholen of scholieren deel kunnen nemen aan een pilot dan dat er belangstelling is, worden doorgaans de meest gemotiveerde deelnemers geselecteerd en vindt er geen randomisatie plaats. Dit leidt doorgaans tot een overschatting van de effectiviteit van de maatregel, als de effecten op langere termijn al geanalyseerd worden.

Om toch effecten van de inrichting van het onderwijs te kunnen meten, zijn onderzoekers op zoek gegaan naar onbedoelde toevallige elementen in het leerproces. Met name Lavy en co-auteurs hebben voor Israel op deze manier zeer interessante analyses gedaan over de optimale inrichting van het onderwijs. Gedacht kan worden aan toevalligheden zoals de geboortemaand van leerlingen, die een verschil veroorzaakt in de precieze leeftijd waarop men met onderwijs begint (Bedard, Dhuey, 2006), de indeling in schooldistricten (Lavy, 2006), migratiegolven (Gould, et al., 2004, Gould, et al., 2006), regels voor de klassengrootte (Angrist en Lavy, 1999), lotterijen die gebruikt zijn om bij een te grote belangstelling computers voor het onderwijs te verdelen (Angrist en Lavy, 2002)

Hoewel het quasi-experimentele en experimentele onderzoek vanuit een statistisch perspectief de betrouwbaarste resultaten geeft, blijken maar heel weinig van de genoemde voorbeelden van dit soort studies betrekking te hebben op de meest fundamentele vragen over de inrichting van het onderwijs. De reden daarvoor is dat onderzoekers die deze aanpak hanteren afhankelijk zijn van de beschikbare data en quasi-experimentele variatie, wat vanuit hun perspectief een grote mate van toevalligheid heeft. Om statistische kwaliteit te kunnen leveren die nodig zijn voor de standaarden van de wetenschappelijke tijdschriften worden deze toevallige omstandigheden bepalend voor de onderzoeksvraag. Niet de afwegingen die het grootste belang hebben voor het onderwijsbeleid, maar de afwegingen die toevallig op een elegante wijze onderzocht kunnen worden, krijgen in dit onderzoeksproces prioriteit.

Vanwege de sterke statistische oriëntatie van dit onderzoek zijn economische analyses over hoe de productie van menselijk kapitaal functioneert nauwelijks ontwikkeld. Alleen op grond van dergelijke analyses zou echter bepaald kunnen worden waar de sleutelparameters zitten die bepalend zijn voor het succes van het onderwijssysteem in zijn geheel. Zo is er een zeer groot aantal statistische analyses uitgevoerd over de effecten van klassenverkleining, maar heeft alleen Lazear (2001) in kaart gebracht waarom effecten van klassenverkleining te verwachten zijn. Op dit moment is het vooral Heckman (Zie bijvoorbeeld Cunha en Heckman, 2006, 2007) die op systematische wijze de technologie van leren in kaart aan het brengen is om de cruciale factoren in het leerproces te achterhalen. Hierbij worden inzichten uit de economische theorie van het menselijk kapitaal, kennis uit de psychologie en schattingen op basis van econometrische modellen gecombineerd om een zo nauwkeurig mogelijk totaalbeeld te krijgen van het leerproces.

Een grondige analyse van de opbouw van het onderwijs in Nederland en de rol die de afzonderlijke onderdelen vervullen bij de realisatie van het eindresultaat is vereist om

vast te kunnen stellen welke aanpassingen in het programma echt een grote bijdrage kunnen leveren aan het eindresultaat. Experimenteel onderzoek van het onderwijssysteem zou zich vervolgens op deze elementen moeten richten, zodat de potentie van onderwijsverbetering en niet de opportuniteit van een wetenschappelijk onderzoek, richtinggevend wordt voor het “evidence based” onderwijsbeleid.

4 Einddoelen, inputs en tussendoelen

Om adequaat het onderwijsproces te kunnen monitoren en aanknopingspunten voor beleid gericht op verbeteringen te vinden, moet het onderwijs dus als proces worden gezien. Hierbij hebben de activiteiten in verschillende levensfasen een specifieke rol binnen de gehele ontwikkeling van het individu.

Omdat in de praktijk de effectiviteit van het onderwijs te kunnen vaststellen, moet in ieder geval duidelijk zijn wat de doelen van dit beleid zijn. Het zou fraai zijn als deze doelen op zijn minst globaal kunnen worden gemeten. De werkdimensie is hierbij doorgaans goed vertegenwoordigd. Als we loon en het hebben van werk beschouwen als doelvariabelen dan is er relatief veel bekend over het arbeidsmarktsucces van mensen. Vaak wordt ook gewezen op andere doeleinden van het onderwijs, zoals algemene ontplooiing, gezondheid, burgerschapszin, etc. Het zal waarschijnlijk niet haalbaar zijn om deze doelen van een relatief gewicht te voorzien, maar voor een heldere discussie en analyse van het onderwijsbeleid lijkt het wel zinvol om niet alleen over arbeidsmarktindicatoren te beschikken als maatschappelijk gezien ook andere doelen aan het onderwijs verbonden worden. Het ontbreken van een bredere verzameling uitkomstvariabelen betekent dat onderzoekers zich puur zullen richten op het arbeidsmarktsucces, en critici er op zullen blijven wijzen dat er veel meer is dan alleen werk en geld, wat onderwijs belangrijk maakt.

Tegenover de doelen staan de inputs in het onderwijsproductieproces. Hierbij zijn vooral van belang:

- de ouders/opvoeders;
- de tijd die leerlingen in het onderwijs doorbrengen;
- de leraren die hiervoor beschikbaar zijn;
- de (financiële) middelen die voor het onderwijs beschikbaar zijn;

Vanuit de traditie dat het beleid vooral de omvang van de beschikbare middelen moet bepalen en het onderwijs zelf het beste weet hoe hiermee zoveel mogelijk leerwinst kan worden behaald zijn veel indicatoren sterk gericht op de omvang van de investering. Relatief weinig is bekend wat er met deze investeringen wordt gedaan.

Zo komt uit de literatuur naar voren dat de invloed van ouders op het studiesucces van kinderen heel erg groot is. De invloed van het ouderlijk huis is zo groot dat in de studies die leerlingen over de tijd volgen wel zeer duidelijke verschillen tussen kinderen waarneembaar zijn bij een vergelijking van sociale milieus, maar deze lijnen nauwelijks van richting veranderen op het moment dat leerlingen het onderwijs betreden. Over wat

de ouders met hun kinderen doen, en hoe deze overdracht gestalte krijgt is heel weinig bekend. Vaak worden indicatoren als opleidingsniveau van de ouders of aantal meters boeken in de kast, gebruikt als proxies voor deze invloed van ouders. De vraag is echter of alle ouders met een vergelijkbare sociaal economische herkomst op eenzelfde manier in hun kinderen investeren. Er zijn aanwijzingen dat Nederlandse ouders – conditioneel op hun inkomen – relatief veel in hun kinderen investeren: de participatiegraad van vrouwen met kinderen is hardnekkig laag; nu vrouwen meer zijn gaan werken door sterke fiscale prikkels lijken mannen met kinderen hun arbeidstijd terug te schroeven tot vaak 32 uur per week; Nederlandse kinderen eten relatief gezond en gaan vroeg naar bed; en vergeleken met het gemiddelde inkomen in Nederland zijn Nederlandse kinderen erg lang⁷. Daarnaast is erg weinig bekend over de tijd die ouders met hun kinderen doorbrengen. Beschikbare indicatoren hebben nu veelal betrekking op theaterbezoek, beschikbaarheid van boeken, en participatie van ouders op school. Waarschijnlijk zijn dit slechts ruwe benaderingen van de ouderlijke investeringen waar het echt om draait. Om de input “ouders” internationaal gezien goed te kunnen vergelijken zouden we moeten weten op welke manieren ouders in hun kinderen investeren.

Voor de tijd die leerlingen in het onderwijs doorbrengen geldt eenzelfde verhaal. Lange tijd zijn alleen de jaren onderwijs meegenomen als indicatoren voor de input van tijd. Hoeveel uren en hoeveel dagen kinderen op school doorbrengen en wat ze in die tijd doen blijft voor een groot deel buiten beeld. De OECD publiceert al enige tijd een indicator voor het aantal uren per jaar, maar deze was lange tijd gebaseerd op het wettelijk minimum in Nederland en niet op de feitelijke praktijk. Over de verdeling van de lestijd is vrijwel niets bekend. Ik heb onlangs een poging gedaan om bijvoorbeeld op basis van lesroosters een beeld te krijgen van de verschuivingen van onderwijsactiviteiten over de jaren, maar het bleek niet goed mogelijk te zijn deze roosters van langer geleden alsnog te achterhalen. Uit een schatting van de feitelijke uren bij een basisschool bleek de praktijk bovendien sterk af te wijken van het officiële lesrooster (Borghans, 2006). Uit analyse van gegevens uit Californië kwam naar voren dat de introductie van kleinere klassen een sterke invloed heeft gehad op de tijdsbesteding van leraren (Borghans, 2005). Van huiswerkactiviteiten is ook weinig bekend. Om goed te kunnen begrijpen wat er gebeurt in het onderwijs en dit internationaal te kunnen vergelijken zou een tijdsbestedingsonderzoek van leerlingen zeer belangrijk zijn. Nederland heeft een sterke traditie op dit front doordat al sinds 1975 elke vijf jaar de tijdsbesteding van Nederlanders wordt gemeten. Hierbij wordt het privéleven van de Nederlander heel precies in kaart gebracht. Wat er gebeurt op het werk en wat er gebeurt op school blijft in dit onderzoek echter buiten beeld.

Over de aantallen leraren en de leerling-leraar-ratio is veel meer bekend. Er is zeer veel, ook zeer degelijk onderzoek gedaan naar de invloed van klassengrootte op prestaties van leerlingen en uit administratieve gegevens en surveys is nauwkeurig bekend hoeveel leraren er zijn en welke achtergrondkenmerken zij hebben. Twee witte vlekken lijken hier

⁷ Lichaamlengte is over het algemeen een zeer nauwkeurige indicator voor welvaart. In dit perspectief zijn Nederlanders sinds enkele decennia echter veel langer dan men op grond van het inkomen zou verwachten (Komlos, Bauer, 2004). De meest logische verklaring zou zijn dat Nederlandse ouders een groter deel van hun welvaart in kinderen investeren dan dat men dat in andere landen doet.

echter ook nog te bestaan. Ten eerste komt uit onderzoek naar voren dat verschillen in kwaliteit tussen leraren – ook al hebben ze dezelfde achtergrondkenmerken – zeer groot kunnen zijn (Zie vooral Rivkin, Hanushek, Kain, 2005). Als alle leraren even goed zouden presteren als de leraren boven in de verdeling zou dit een enorme vergroting van het rendement van onderwijs met zich meebrengen. Het blijkt echter zeer moeilijk om te achterhalen welke kenmerken deze leraren zo verschillend maken. Daarnaast bestaat een hardnekkig gerucht dat de leraren van tegenwoordig niet meer de leraren van vroeger zijn. In Nederland heeft er in ieder geval bij de leraren op de middelbare school een verschuiving plaatsgevonden van universitair opgeleiden en leraren met een MO-acte naar leraren met een NLO-opleiding. Uit het onderzoek van Rivkin et al. komt echter naar voren dat de formele opleidingsachtergrond van leraren weinig zegt over de kwaliteit. Interessanter is daarom om te bekijken hoe het feitelijk niveau van de leraren zich heeft ontwikkeld. De Nederlandse bevolking loopt zeer uiteen qua talenten. De vraag is welke segment uit deze verdeling op de Nederlandse arbeidsmarkt wordt geselecteerd voor het leraarschap en hoe dit selectieproces zich in de loop der tijd heeft ontwikkeld.

Tussendoelen

In principe zou voor een analyse van het onderwijs voldoende moeten zijn om inputs aan resultaten te relateren. Het gaat echter vaak om effecten op zeer lange termijn, terwijl bovendien niet alleen de kwaliteit van het onderwijs bepalend is voor die doelen die men aan onderwijs zou kunnen stellen. Voor werk geldt bijvoorbeeld dat ook institutionele kenmerken van de arbeidsmarkt in hoge mate bepalend zijn voor de uitkomsten. In Frankrijk is bijvoorbeeld de werkloosheid onder schoolverlaters zeer hoog in vergelijking met andere landen. Het is niet plausibel dat dit een gevolg is van het Franse onderwijs.

Om op een directere manier de effectiviteit van de inrichting van het onderwijs te kunnen vaststellen is het daarom belangrijk om ook prestatie-indicatoren tijdens de onderwijsloopbaan vast te stellen. Leren stopt bovendien niet op de dag dat mensen het voltijds onderwijs verlaten. Het is daarom ook zinvol het leerproces en de vaardigheden van mensen die het onderwijs reeds hebben verlaten te volgen.

Dit is de essentie van de internationaal vergelijkende testen in PIRLS, TIMSS, en PISA. Een belangrijke vraag is echter welke vakgebieden moeten worden gemeten, wanneer dit het beste kan gebeuren en wat de betekenis is van de schaal die hierbij wordt gehanteerd.

De initiatieven voor internationaal vergelijkende testen waren in eerste instantie vooral gericht op de eigen taal en wiskunde/natuur en techniek. Op school leren kinderen echter veel meer vakken, waarvan kennelijk wordt aangenomen dat ze waardevol zijn. Leung (2002) heeft de hoge scores bij TIMSS van Singapore geanalyseerd en concludeert dat de goede wiskundeprestaties ten koste zijn gegaan van andere vakken. Scores op de vakken die getest worden zullen daarom zowel afhangen van de aandacht die tot dan toe is gegeven aan deze vakken en de kwaliteit/effectiviteit van dit onderricht. Tijdens de middelbare-schoolperiode is dit probleem nog redelijk overzichtelijk. Vrijwel iedereen zal het er mee eens zijn dat taal en wiskunde in ieder geval belangrijke onderdelen van het lesprogramma zijn. Voor het tertiaire onderwijs en vooral bij het meten van kennis en

vaardigheden van werkenden is dit veel minder het geval. De IALS heeft interessante informatie opgeleverd over de geletterdheid van de bevolking, maar als maat voor de beroepsrelevante kennis- en vaardigheden schiet dit natuurlijk tekort. Bij de opvolger van IALS wordt geprobeerd om een breder scala aan kennis- en vaardigheden te testen maar het fundamentele probleem is dat de kennis die relevant is voor het functioneren van mensen afhangt van hun beroep en dus niet voor de bevolking in de breedte kan worden onderzocht. Alleen al bij het meten van computervaardigheden stuitte men op problemen, doordat testen af zouden hangen van de specifieke computerpakketten waar mensen ervaring mee hebben. In plaats daarvan gebruikt men indicatoren over het gebruik van ICT. Aanvankelijk gingen economen er nog van uit dat het gebruik van ICT een goede indicator is voor de computervaardigheden van mensen (Krueger, 1993), maar inmiddels is duidelijk dat dit niet het geval is (Borghans en Ter Weel, 2004). In de praktijk zijn veel vaardigheden daarom niet te toetsen door middel van een test en moet gebruik worden gemaakt van self-assessment. Binnen de OECD wordt nog gediscussieerd over het meten van vaardigheden van werkenden. Francis Green pleit voor een systeem dat gebaseerd is op de activiteiten die mensen op het werk uitvoeren. Rolf van der Velden pleit voor een systeem waarbij mensen hun vermogen zelf in moeten schatten. De eerste aanpak oogt objectiever maar leunt sterk op de veronderstelling dat de taken op het werk een op een samenhangen met de vaardigheden waarover mensen beschikken. Borghans, Golsteyn en De Grip (2006) hebben methodes ontwikkeld waarmee de kennisaccumulatie van werkenden in het algemeen in kaart kan worden gebracht. Deze methode moet nog verder onderzocht worden, maar interessant is al wel dat naar voren komt dat mensen de eerste jaren op de arbeidsmarkt heel veel leren. Als leren op het werk een substituut is voor leren op school, zouden testprogramma's die zich richten op school (zoals PISA) dus eigenlijk ook de eerste jaren op de arbeidsmarkt moeten afdekken.

Nu richten de vergelijkende studies zich op specifieke leeftijden. De keuze hiervan wordt sterk bepaald door de randvoorwaarde dat in alle landen die deelnemen in principe alle leerlingen nog op school moeten zitten. Als leerlingen het onderwijs al hebben verlaten en toch scholen als steekproefkader worden gebruikt, leidt een hoge drop-out rate tot een hoge score van een land. De consequentie is wel dat de verschillen in scores tussen landen niet alleen afhangen van hoe belangrijk dit vak wordt gevonden en hoe effectief het onderwijs in dit vak is, maar ook van de timing van dit vak in het curriculum. Als het het perspectief van onderwijsproductie gunstig is om laat in het curriculum veel aandacht aan wiskunde of natuur en techniek te besteden, zullen landen die deze optimale strategie volgen slechter scoren in de vergelijkende test.

Vervolgens is de schaal die gehanteerd wordt om de kennis en vaardigheden van leerlingen te vergelijken niet zonder problemen. Bij de totstandkoming van de testen wordt zeer zorgvuldig een keuze gemaakt van de vragen die worden voorgelegd aan de leerlingen en wordt ook veel aandacht besteed aan de statistische kwaliteit. Ondanks deze zorgvuldigheid heeft een testscore altijd een willekeurige schaling. Men kan bij cijfers voor een test op een in Nederland gebruikelijke schaal van 0-10 niet zeggen dat het verschil tussen 2 en 3 even groot is als het verschil tussen 6 en 7 of 9 en 10. Op dezelfde wijze hoeft het verschil tussen 480 en 490 op de internationale schalen niet even groot te zijn als het verschil tussen 550 en 560. Iedere monotone transformatie van een testscore

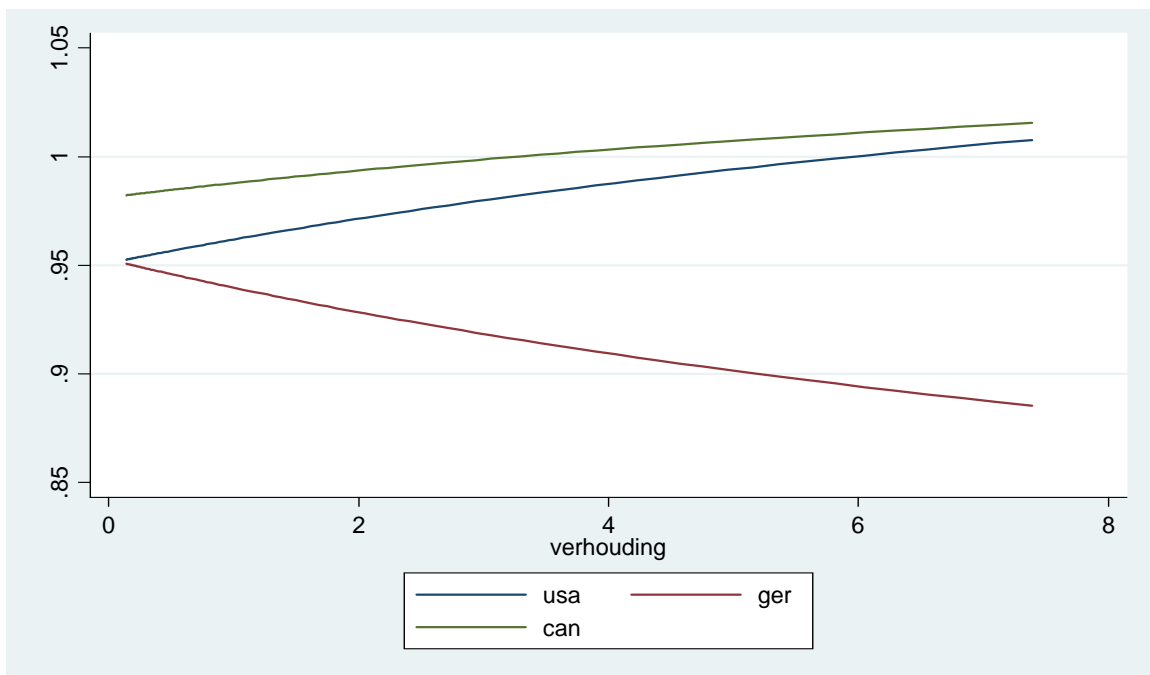
is zelf ook weer een adequate test score. Dit oogt wellicht als een technisch details, maar heeft in de praktijk enorme consequenties. Cunha en Heckman (2006) hebben onlangs laten zien dat een analyse over de vorming van menselijk kapitaal die rekening houdt met deze rekbaarheid van toetsschalen tot fundamenteel andere conclusies leidt dan een analyse die de verhoudingen op de schaal als vaststaand beschouwd.

Met een voorbeeld kan dit effect gemakkelijk getoond worden. In figuur 1 staan de gemiddelde scores voor literacy gemeten door IALS voor werkenden in de VS, Duitsland en Canada in vergelijking met Nederland weergegeven. Om deze gemiddeldes te berekenen is de schaal boven 285 (ongeveer het gemiddelde) uitgerekt of ingekrompen ten opzichte van de schaal beneden 285. Een verhouding van 2 op de x-as betekent dus dat een verschil van hoge scores twee keer zo belangrijk wordt gevonden als in de oorspronkelijke schaal vergeleken met de lage scores. Bij de verhouding 0.5 zijn de lage scores juist twee maal zo belangrijk.

De Nederlandse score fungeert als ijkpunt en is steeds op 1 gezet. Als lage scores relatief belangrijk zijn blijkt Nederland van deze vier landen de koppositie in te nemen. Canada volgt als tweede, Duitsland en de VS eindigen vrijwel gelijk als derde en vierde. Naarmate echter juist de verschillen in de hoge scores relatief zwaarder worden meegeteld, stijgen de VS en Canada ten opzichte van Nederland en daalt de positie van Duitsland verder. In dit voorbeeld eindigt Canada uiteindelijk als 1, de VS als 2 en Nederland als derde.

Figuur 1

Score op de International Adult Literacy Survey van de Verenigde Staten (usa), Duitsland (ger) en Canada (can) in vergelijking met Nederland, waarbij het bovenste deel van de schaal is uitgerekt of ingekrompen ten opzichte van het onderste deel.



De vraag is dan wat de redenen zouden kunnen zijn om een sterk vervormde schaal te prefereren boven de schaal die door de OECD is gehanteerd. Dat hangt af van zowel de moeilijkheidsverdeling van de toetsvragen als de betekenis van de betreffende vaardigheid, bijvoorbeeld op de arbeidsmarkt. In het algemeen bevat een toets veel vragen in een specifieke range van moeilijkheid. Verschillen in het niveau van mensen ver onder of ver boven deze range worden niet meer goed vastgesteld. De testen die gebruikt worden voor de internationale vergelijkingen zijn sterk gericht op de lagere kennisniveau's. Verschillen bij de betere leerlingen komen dus minder sterk tot uitdrukking in de score.

Daarnaast is echter ook de vraag wat een testscore waard is in termen van de einddoelstelling van het onderwijs. De redenering is het makkelijkste op te zetten voor loon, maar geldt voor iedere doelvariabele. Bij geletterdheid zou men zich kunnen voorstellen dat vooral hele lage niveau's zeer negatief werken voor de productiviteit van werkenden. In dat geval zou juist de onderkant van de schaal opgerekt moeten worden. Bij wiskunde kan men zich voorstellen dat juist de verschillen tussen de toppers van groot belang zijn. Dat zullen immers de personen zijn die deze vaardigheden ook in gaan zetten op de arbeidsmarkt en dus het verschil gaan uitmaken qua economische performance. De betekenis van de testcores zal dus afhangen van de manier waarop deze vaardigheden worden benut op de arbeidsmarkt. Om adequaat de cijfers te kunnen interpreteren zou deze link tussen tussenscore en einddoelen duidelijk moeten zijn, maar in de praktijk is hierover weinig bekend. De OECD en IEA zijn zich bewust van het arbitraire karakter van de schaling van de testcores en rapporteren daarom bij voorkeur een overzicht waarin de hele verdeling wordt weergegeven. In de praktijk hebben de gemiddelde scores echter vooral invloed op het beleid en wordt in onderzoek ook bij gebrek aan alternatieven dit gemiddelde als maatstaf gebruikt. Hoewel de officiële rapporten de nadruk op de gehele verdeling leggen, geven ze geen informatie over de betekenis van de punten op de schaal voor bepaalde einddoelen van het onderwijs.

Naast dit schalingsprobleem is het ook niet duidelijk of de condities waaronder de testen in de verschillende landen worden afgenomen wel gelijk zijn. Bij het invullen van een TIMSS- of PISA-test staat er voor de leerlingen zelf niets op het spel, maar is wellicht wel de eer van het land in het geding. Uit analyses van Borghans, Meijers en Ter Weel (2007) en Segal (2007) komt naar voren dat studenten verschillend reageren op testvragen waarbij geen incentives aanwezig zijn. Studenten met bepaalde persoonseigenschappen hebben de neiging om altijd hun best te doen, andere doen met name hun best als er veel op het spel staat. Als leerlingen in verschillende landen verschillen qua persoonskenmerken of een sterkere neiging hebben hun best te doen voor de eer van hun land, kunnen de scores beïnvloedt worden. Het vermoeden bestaat dat dergelijke effecten deels de opmerkelijke verschillen tussen Westerse en Aziatische landen kunnen verklaren.

5 Conclusie

Hoewel er veel interesse bestaat voor internationaal vergelijkende gegevens over het onderwijs, blijkt dat de invloed die deze gegevens hebben gehad op de inrichting van het onderwijs in Nederland beperkt is. Er bestaan grote verschillen in de manier waarop landen hun onderwijs vormgeven. Het is daarom opmerkelijk dat nog steeds zeer weinig bekend is over de consequenties van deze keuzes over de inrichting van het onderwijs voor de eindresultaten van het onderwijs.

De indicatoren zoals die te vinden zijn in Education at a Glance zijn zeker zinvol om het Nederlandse onderwijs in perspectief te plaatsen. Omdat ze buiten de context van het Nederlandse onderwijs tot stand komen bieden ze ook een vorm van objectiviteit en bovenal doordat de gegevens van de internationaal testen vrij beschikbaar zijn, bieden ze vele onderzoekers een interessante bron van gegevens. Ook in het “ijken” van verschillende niveau’s van onderwijs hebben deze gegevens een belangrijke rol gespeeld.

Door het onderwijs te beschouwen als een proces waarbij alle onderdelen een bijdrage leveren aan de eindresultaten wordt duidelijk dat de betekenis van internationale indicatoren voor onderzoek naar mogelijke verbeteringen van het onderwijssysteem beperkt zal blijven. Indicatoren zijn veelal tussenstanden in dit onderwijsproces. Doordat landen in sterke mate verschillen in de wijze waarop ze het onderwijs hebben ingericht, kunnen deze tussenstanden ook alleen goed begrepen worden als alle details van het systeem hierin zijn meegewogen. In de praktijk is het daarom veel makkelijker om variaties die zich voordoen binnen een land te gebruiken voor dergelijke analyses. Statistische gezien leveren dergelijke (quasi-)experimentele onderzoeken veel betrouwbaardere uitkomsten op. Vrijwel alle succesvolle wetenschappelijke analyses van het onderwijs hebben dan ook betrekking op gegevens over een specifiek land.

Toch lijkt het ook moeilijk te zijn om goed vorm te geven aan dit “evidence based” onderwijsbeleid. De oorzaak hiervan is dat het bestaande onderzoek in deze traditie sterk afhankelijk is van toevallige mogelijkheden die de data biedt en daardoor de data en niet het belang van bepaalde vragen of het onderwijs, bepalend wordt voor de probleemstelling van het onderzoek.

Om er voor te zorgen dat het onderzoek zich kan richten op de meest cruciale afwegingen die in het onderwijs moeten worden gemaakt, zouden allereerst gegevens beschikbaar moeten komen die het mogelijk maken de ontwikkeling van kennis en vaardigheden bij mensen over hun levensloop in beeld te brengen. Hierbij is nog weinig bekend over de rol van ouders in de ontwikkeling van kinderen, de tijdsbesteding op school (en aan leren thuis), en is ook weinig bekend over de ontwikkeling van non-cognitieve vaardigheden. Daarnaast is het van belang om de metingen van de vaardigheden van leerlingen te verbreden: Om de ontwikkeling van mensen goed in kaart te kunnen brengen zouden testen op meerdere momenten in en na de schoolloopbaan afgenomen moeten worden en zou het ook waardevol zijn als deze op meer vakgebieden betrekking hadden dan alleen taal en wiskunde. Een longitudinaal karakter zou de vergelijking tussen ontwikkeling en latere uitkomsten mogelijk maken.

Hoewel de internationaal vergelijkende dimensie zeker interessant is om zicht te krijgen op de vraag waar de ontwikkeling van Nederlandse kinderen afwijkt van leerlingen in andere landen en hoe dit samenhangt met de inrichting van het leerproces, is het de vraag of het reëel is dergelijke informatie in internationaal verband te verzamelen. Dit vergt veel coördinatie, terwijl er bovendien in Nederland mogelijkheden voor dataverzameling zijn die in andere landen niet beschikbaar zijn.

Recente ontwikkelingen bij de dataverzameling van het CBS maken het mogelijk gegevens uit verschillende fasen van het onderwijs op persoonsniveau aan elkaar te relateren. Bekeken zou moeten worden welke gegevens hiervoor via administratieve bestanden in principe beschikbaar zijn en welke gegevens hiervoor van belang zijn. In aanvulling hierop zouden op steekproefbasis extra gegevens verzameld kunnen worden. Een voorbeeld van een dergelijke aanpak is de koppeling van de VOCL-cohorten aan de salarisgegevens van het Sociaal Statistisch Bestand waardoor de deelnemers aan het cohort over een veel langere periode kunnen worden gevolgd.

Omdat onderwijs een langdurig proces is zijn de gegevens die nu bewaard gaan worden de bron van kennis voor over 20 jaar. Het is daarom van groot belang om grondig te inventariseren welke gegevens bewaarde zouden moeten worden, zodat deze beslissingen niet op ad hoc basis genomen zullen worden.

Daarnaast betekent het bestaan van gegevens nog niet dat elke onderzoeker hier gebruik van kan maken. Tot voor kort moesten onderzoekers nog aanzienlijke bedragen betalen om gebruik te mogen maken van bijvoorbeeld PRIMA of VOCL. Gegevens zoals PPOZ zijn voor onderzoekers ook niet goed toegankelijk. Als de overheid wil bevorderen dat onderzoekers zich richten op maatschappelijk relevante onderzoeksvragen, is zorgen dat de juiste onderzoeksdata beschikbaar zijn de beste methode. Door een groot aantal gegevens die nu in het kader van het onderwijsproces worden verzameld, centraal op te slaan en te relateren aan het SSB van het CBS en er voor te zorgen dat onderzoekers deze gegevens ook daadwerkelijk kunnen gebruiken, kan een grote vooruitgang geboekt worden in het evidence based onderzoek dat werkelijk bijdraagt aan het verbeteren van het Nederlandse onderwijssysteem.

Literatuur

Antenbrink, Patrick, Kok Burger, Maarten Cornet, Marieke Rensman & Dinand Webbink (2005), *Nederlands onderwijs en onderzoek in internationaal perspectief*, CPB document 88.

Ammermuller, Andreas, Hans Heijke & Ludger Wossmann (2005), "Schooling Quality in Eastern Europe: Educational Production during Transition." *Economics of Education Review* 24(5): pp. 579-599.

Ammermueller, Andreas & Jorn-Steffen Pischke (2006), "Peer Effects in European Primary Schools: Evidence from PIRLS." NBER Working Papers: 12180.

Angrist, Joshua D. & Victor Lavy (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114(2): pp. 533-575.

Angrist, Joshua D. & Victor Lavy (2002), "New Evidence on Classroom Computers and Pupil Learning." *Economic Journal* 112(482): pp. 735-765.

Barro, Robert J. & Jong-Wha Lee (1993), "International Comparisons of Educational Attainment." *Journal of Monetary Economics* 32(3), pp. 363-394.

Barro, Robert J. & Jong-Wha Lee (2001), "International Data on Educational Attainment: Updates and Implications." *Oxford Economic Papers* 53(3): pp. 541-563.

Bedard, Kelly & Elizabeth Dhuey (2006), "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects." *Quarterly Journal of Economics* 121(4): pp. 1437-1472.

Blau, Francine D. & Lawrence M. Kahn (1996), "International Differences in Male Wage Inequality: Institutions versus Market Forces." *Journal of Political Economy* 104(4): pp. 791-836.

Borghans, Lex (2006), "'Zonde van de tijd' Leren in Nederland vanuit een economisch perspectief." Oratie Maastricht.

Borghans, Lex (2005), "Class Teaching and Individual Instruction."

Borghans, Lex, Angela Duckworth, James Heckman & Bas ter Weel (2007), "The Economics of Non-Cognitive Skills."

Borghans, Lex, Bart Golsteyn & Andries de Grip (2006), "Meer Werken is meer leren: De Determinanten van Kennisontwikkeling", Centrum voor Innovatie van Opleidingen, Den Bosch.

Borghans, Lex, Francis Green & Ken Mayhew (2001), "Skills Measurement and Economic Analysis: An Introduction." *Oxford Economic Papers* 53(3): pp. 375-384.

Borghans, Lex & Hans Heijke (2005), "The Production and Use of Human Capital: Introduction." *Education Economics* 13(2): pp. 133-142.

Borghans, Lex, Huub Meijers & Bas ter Weel (2007), "The Role of Noncognitive Skills in Explaining Cognitive Test Scores." *Economic Inquiry*, forthcoming.

Borghans, Lex & Bas ter Weel (2004), "Are Computer Skills the New Basic Skills? The Returns to Computer, Writing and Math Skills in Britain." *Labour Economics* 11(1): pp. 85-98.

Cornet, M., F. Huizinga, B. Minne & D. Webbink (2006), “Kansrijk kennisbeleid.” CPB Den Haag.

Cunha, Flavio & James J. Heckman (2006), “Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation.” Working paper Chicago.

Cunha, Flavio & James J. Heckman (2007). “The technology of skill formation.” NBER Working paper.

Freeman, Richard & Ronald Schettkat (2001), “Skill Compression, Wage Differentials, and Employment: Germany vs the US.” *Oxford Economic Papers* 53(3): pp. 582-603.

Gollop, F.M. & D.W. Jorgenson (1980), “US Productivity by Industry, 1947-1973.” In J. Kendrick en B. Vaccara (eds.), *New Developments in Productivity Measurement*, Chicago.

Gould, Eric D., Victor Lavy & Daniele M. Paserman (2004), “Immigrating to Opportunity: Estimating the Effect of School Quality Using a Natural Experiment on Ethiopians in Israel.” *Quarterly Journal of Economics* 119(2): pp. 489-526.

Gould, Eric D., Victor Lavy & Daniele M. Paserman (2006), “Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence.” C.E.P.R. Discussion Papers 5439.

Hanushek, Eric A. & Ludger Wossmann (2006), “Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries.” *Economic Journal* 116(510): pp. C63-76

Hanushek, Eric A. & Javier A. Luque (2003), “Efficiency and Equity in Schools around the World.” *Economics of Education Review* 22(5): pp. 481-502.

Hanushek, Eric A. (2003), “The Failure of Input-Based Schooling Policies.” *Economic Journal* 113(485), pp. F64-98.

Hoxby, Caroline (2000), “Does Competition among Public Schools Benefit Students and Taxpayers?” *American Economic Review* 90(5): pp. 1209-1238.

Hoxby, Caroline (2004), “Productivity in Education: The Quintessential Upstream Industry.” *Southern Economic Journal* 71(2): pp. 209-231.

Hoxby, Caroline (2005), “Competition Among Schools: A Reply to Rothstein (2004).” NBER Working Papers: 11216.

Jacob, Brian A. & Lars Lefgren (2005), "What Do Parents Value in Education? An Empirical Investigation of Parents' Revealed Preferences for Teachers." NBER Working Papers: 11494.

Jacob, Brian A. & Steven D. Levitt (2003), "Catching Cheating Teachers: The Results of an Unusual Experiment in Implementing Theory." *Brookings-Wharton Papers on Urban Affairs*, pp. 185-209

Jurges, Hendrik & Kerstin Schneider (2004), "International Differences in Student Achievement: An Economic Perspective." *German Economic Review* 5(3): pp. 357-380.

Komlos, John & Marieluise Baur (2004), "From the Tallest to (One of) the Fattest: The Enigmatic Fate of the American Population in the 20th Century." *Economics and Human Biology* 2(1): pp. 57-74.

Krueger, Alan B. (1993), "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989." *Quarterly Journal of Economics* 108(1), pp. 33-60.

Lavy, Victor (2004), "Performance Pay and Teachers' Effort, Productivity and Grading Ethics." NBER Working Papers: 10622.

Lavy, Victor (2006), "From Forced Busing to Free Choice in Public Schools: Quasi-Experimental Evidence of Individual and General Effects." NBER Working Papers: 11969.

Lazear, Edward P. (2001), "Educational Production." *Quarterly Journal of Economics*, 116(3): pp. 777-803.

Leung, Frederick (2002), "Behind the High Achievement of East Asian Students." *Educational Research and Evaluation* 8(1): pp. 87-108.

Leuven, Edwin, Hessel Oosterbeek & Hans van Ophem (2004), "Explaining International Differences in Male Skill Wage Differentials by Differences in Demand and Supply of Skill." *Economic Journal* 114(495): pp. 466-86.

Martin, M.O., I.V.S. Mullis, E.J. Gonzalez & S.J. Chrostowski (2004), "Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades." Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Medrich, Elliott & Jeanne Griffith (1992), "International Mathematics and Science Assessment: What Have We Learned?" National Center for Education Statistics Research and Development Report 011.

Mullis, I.V.S., M.O. Martin, E.J. Gonzalez & A.M. Kennedy (2003), "PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools." Chestnut Hill, MA: Boston College.

Mullis, I.V.S., M.O. Martin, E.J. Gonzalez & S.J. Chrostowski (2004), "Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades." Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Murnane, Richard & Richard Nelson (2005), "Improving the Performance of the Education Sector: The Valuable, Challenging, and Limited Role of Random Assignment Evaluations." NBER working paper 11846.

OECD, Statistics Canada (2001), "Literacy in the Information Age Final Report of the International Adult Literacy Survey."

OECD (2001), "Knowledge and Skills for Life: First Results from PISA 2000."

OECD (2004), "Learning for Tomorrow's World – First Results from PISA 2003."

OECD (2006) "Education at a Glance 2006."

Onderwijsraad (2006), "Naar meer evidence based onderwijs."

Onderwijsraad (2005), "De stand van educatief Nederland."

Rivkin, Steven G., Eric A. Hanushek & John F. Kain (2005), "Teachers, Schools, and Academic Achievement." *Econometrica* 73(2): pp. 417-58.

Rothstein, Jesse (2005), "Does Competition Among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby (2000)." NBER Working Papers: 11215.

SCP (2007), *Publieke prestaties in perspectief*, Sociaal Cultureel Planbureau.

Segal, Carmit (2006), "Motivation, Test Scores, and Economic Success"

Steedman, Hilary & Steven McIntosh (2001), "Measuring Low Skills in Europe: How Useful Is the ISCED Framework?" *Oxford Economic Papers* 53(3): pp. 564-581.

Urquiola, Miguel & Chang-Tai Hsieh (2002) "When Schools Compete, How Do They Compete? An Assessment of Chile's Nationwide School Voucher Program."

Wossmann, Ludger & Martin West (2006), "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS." *European Economic Review* 50(3): pp. 695-736.

Wossmann, Ludger (2005), "The Effect Heterogeneity of Central Examinations: Evidence from TIMSS, TIMSS-Repeat and PISA." *Education Economics* 13(2): pp. 143-169.