

Jan van de Craats

Korteweg-de Vries Instituut

Science Park 904

1098 XH AMSTERDAM

J.vandeCraats@uva.nl

Onderwijs Meten van kwaliteit

TIMSS 2007 – Feiten en meningen

De kwaliteit van het rekenonderwijs op de basisschool blijft een twistpunt. Het internationale onderzoek TIMSS dat deze kwaliteit voor verschillende landen op een rijtje zet, wordt door verschillende partijen gebruikt om te illustreren hoe goed dan wel slecht het onderwijs ervoor staat. Elders in dit nummer reageren Marja van den Heuvel-Panhuizen en Nellie Verhoef op het redactioneel van het vorige NAW-nummer. In onderstaand artikel licht Jan van de Craats zijn bezwaren tegen een positieve interpretatie van het TIMSS 2007 rapport toe.

TIMSS staat voor *Trends in International Mathematics and Science Study*, een periodiek georganiseerd internationaal vergelijkend onderzoek naar prestaties van leerlingen op het gebied van de wiskunde en de natuurwetenschappen. Het laatste TIMSS onderzoek waarover gerapporteerd is, is gehouden in 2007. Er is toen onderzoek uitgevoerd op twee gebieden, *mathematics* en *science* in twee leeftijds-categorieën: *grade 4*, overeenkomend met onze groep 6 van het basisonderwijs, en *grade 8*, overeenkomend met klas 2 van het voortgezet onderwijs. In 2007 heeft Nederland alleen meegedaan aan het onderzoek in groep 6 van de basisschool.

Ik beperk me in dit artikel tot het deel *mathematics* (rekenen en wiskunde) van het TIMSS-onderzoek voor *grade 4*. In totaal hebben daaraan in 2007 zesendertig landen meegedaan. De uitkomsten van TIMSS 2007 op het gebied *mathematics* van de Nederlandse deelnemers worden in bepaalde kringen geïnterpreteerd als ondersteuning van de stelling dat het Nederlandse rekenonderwijs op de basisschool internationaal gezien van hoog niveau is. Daarbij wordt soms de kreet ‘*best of the west*’ gehanteerd. Ik zal laten zien dat de resultaten van TIMSS 2007 een dergelijke conclusie niet toelaten. Die juichtoon is ook niet in lijn met de inhoud van het geza-

menlijke persbericht van NWO en de Universiteit Twente over TIMSS 2007 (9 december 2008) [5].

Feiten

De volgende feiten zijn ontleend aan de officiële rapportages [1–4] die ook via het internet geraadpleegd en gedownload kunnen worden.

1. In de landenrangschikking eindigde Nederland op de negende plaats, met de aantekening dat Nederland niet voldeaan heeft aan de TIMSS-richtlijnen voor deelname via een representatieve steekproef (p. 34 in [3]). Daardoor mogen de resultaten van TIMSS 2007 *niet* als representatief worden beschouwd voor de gehele populatie van Nederlandse basisschoolleerlingen van groep 6.
2. Dat Nederland niet voldeaan heeft aan de TIMSS-richtlijnen kan als volgt nader worden gespecificeerd (pp. 178–180 in [4], pp. 12–16 in [1]). Net zoals in de andere deelnemende landen was er ook in Nederland een blinde steekproef van basisscholen getrokken. In Nederland omvatte die steekproef 150 scholen. In elk van die scholen moest minstens één willekeurig getrokken volledige klas in groep 6 meedoen. Twee scholen vielen na de trekking

af omdat ze inmiddels waren opgeheven.

Bij al dit soort onderzoeken is het zo, dat alleen als een voldoende hoog percentage uit de steekproef ook daadwerkelijk meedoet, de resultaten van het onderzoek representatief genoemd mogen worden. Volgens de TIMSS-richtlijnen zou minstens 85 procent van de scholen in de steekproef ook daadwerkelijk moeten meedoen (pp. 386, 392 in [3], p. 13 in [1]). Zeven landen voldeden niet aan dit criterium. Zes van de zeven landen haalden wel een deelnamepercentage van minstens 70 procent. Eén land voldeed echter bij lange na niet aan de eis. Dat was Nederland, met een deelnamepercentage van slechts 48 procent (72 van de 150 scholen, (p. 180 in [4]).

3. Bij de zeven landen (waaronder Nederland) die niet aan het TIMSS-criterium voldeden, zijn weigerende scholen uit de steekproef vervangen door andere scholen. In Nederland waren er 69 vervangende scholen, zodat er in totaal 141 scholen hebben meegedaan. Het totale aantal Nederlandse leerlingen dat mee heeft gedaan, bedraagt 3349, gemiddeld bijna 24 per school (p. 182 in [4]). Daarmee kwam het aantal deelnemers weer op een redelijke hoogte. Deze handelwijze heeft echter tot gevolg dat de resultaten dan niet meer representatief genoemd mogen worden. Ze zeggen dan alleen nog maar iets over de *deelnemerspopulatie* (de groep van alle deelnemers aan het TIMSS-onderzoek in Nederland), maar nauwelijks iets over de *totale populatie* van alle Nederlandse scholieren in groep 6. De TIMSS-



rapportage heeft de resultaten van de landen die gebruik hebben gemaakt van vervangende scholen dan ook in de tabellen van een speciaal merkteken voorzien om dit aan te geven.

4. In de Nederlandse deelnemerspopulatie presteerden meisjes significant slechter dan jongens. Allochtone leerlingen presteerden significant slechter dan autochtone leerlingen. Allochtone meisjes presteerden significant slechter dan allochtone jongens (pp. 27–28 in [1]).
5. Onderzocht is ook de *positieve houding* van de deelnemers tegenover rekenen en wiskunde. De resultaten van dat onderzoek staan vermeld in een tabel op bladzijde 175 van [3]. In deze tabel van zesendertig landen staat de Nederlandse deelnemersploeg op de *vierendertigste* plaats. De Nederlandse deelnemers hebben dus heel weinig waardering voor rekenen en wiskunde.
6. Als je de resultaten van TIMSS 2007 ver-

gelijkt met die van de vorige lichtingen (TIMSS 2003 en TIMSS 1995) valt op dat Nederland behoort tot de weinige landen waarbij er een *gestage achteruitgang* te zien is in de rekscores (pp. 44–45 in [3], pp. 25–26 in [1]).

7. TIMSS 2007 *grade 4*, toetst uitsluitend leerlingen in groep 6 van het basisonderwijs. Hoe de resultaten van het rekenonderwijs *aan het einde* van de basisschool zijn, wordt in TIMSS dus niet getoetst. Juist over dat eindniveau zijn er in Nederland veel klachten.

In Tabel 1 staan de landenscores van de beste 20 landen in TIMSS 2007. Ze zijn voorzien van een sterretje wanneer de deelname in de steekproef minder dan 85 procent was. Ook zijn de scores van TIMSS 2003 vermeld indien die landen toen ook hebben meegedaan.

Meningen

Voorstanders van realistisch rekenen gebruiken de TIMSS resultaten om te betogen dat-

het Nederlandse rekenonderwijs het internationaal gezien uitstekend doet. Dit is niet gerechtvaardigd vanwege de te geringe respons op de aselechte steekproef.

Afgezien hiervan is het niet duidelijk waarom de voorstanders van realistisch rekenen zo nadrukkelijk de resultaten van de niet-westerse landen die hoog eindigen, in hun beoordeling buiten beschouwing laten (Nederland zou *'best of the west'* zijn). Wiskunde en rekenen kennen geen grenzen. Als Singapore, Taiwan en Japan verreweg de beste resultaten halen, dan valt er blijkbaar iets te leren van hun rekenonderwijs. En, daarover is geen twijfel, dat onderwijs is niet realistisch maar traditioneel.

De TIMSS-resultaten voor de Nederlandse deelnemers zijn niet bemoedigend, maar juist uitermate verontrustend: de *lage positieve attitude* van de Nederlandse deelnemers spreekt boekdelen. (Voor veel Nederlandse basisschooldocenten geldt overigens hetzelfde.) Realistisch rekenen heeft de pretentie re-

	2007	2003
Hong Kong	607*	575*
Singapore	599	594
Chin. Taipei	576	564
Japan	568	565
Kazachstan	549	
Rusland	544	532
Engeland	541*	531*
Letland	537	536
Nederland	535*	540*
Litouwen	530	534
USA	529*	518*
Duitsland	525	
Denemarken	523*	
Australië	516	499*
Hongarije	510	529
Italië	507	503
Oostenrijk	505	
Zweden	503	
Slovenië	502	479
Armenië	500	456

Tabel 1 Scores van TIMSS 2007

kenen leuk te maken voor kinderen. Blijkens TIMSS is bij de Nederlandse deelnemers juist het tegendeel het geval.

Bij de West-Europese deelnemers valt op dat er enige belangrijke landen niet hebben deelgenomen: Spanje, Frankrijk, Finland, België (Vlaanderen), om er enige te noemen. Ook dat maakt de claim 'best of the west' twijfelachtig.

Een beoordeling van de waarde en de relevantie van de TIMSS-onderzoeken wordt bemoeilijkt doordat de helft van de gebruikte opgaven geheim wordt gehouden. De officiële verklaring is dat men die opgaven in een later onderzoek weer wil gebruiken.

Een gedachtenexperiment

Dat een hoge respons op de originele steekproef (zonder vervangingsscholen) noodzakelijk is om statistisch verantwoorde conclusies te trekken zal intuïtief misschien wel duidelijk zijn, maar het is de moeite waard om dit toch wat nader toe te lichten. Ik doe dat in de vorm van een gedachtenexperiment waarin ik de werkelijkheid wat vereenvoudig.

Stel dat aan elke school in Nederland een *rekenvaardigheidsindex* r kan worden toegekend, een getal tussen 1 en 10 dat aangeeft

hoe goed de kinderen in groep 6 op die school rekenen. Stel, om de gedachten te bepalen, dat de waarden van r verdeeld zijn rond een gemiddelde van $5\frac{1}{2}$. Ook in de steekproef van 148 scholen mag je dan ongeveer zo'n verdeling met een gemiddelde van $5\frac{1}{2}$ verwachten. Stel nu dat de bereidheid om aan het onderzoek deel te nemen positief gecorreleerd is met de rekenvaardigheidsindex: scholen die goed zijn in rekenen doen graag mee, scholen die niet goed zijn zullen bij voorkeur weigeren. Dan is de gemiddelde rekenvaardigheidsindex bij de 72 scholen die meedoen aanzienlijk groter dan $5\frac{1}{2}$.

Er waren $2 \times 78 = 156$ reservescholen (voor elke weigerende of afgevalen school twee), waarvan er uiteindelijk 69 ook hebben meegedaan. Laten we voor het gemak aannemen dat de gemiddelde rekenvaardigheidsindex onder die 69 vervangingsscholen weinig afwijkt van $5\frac{1}{2}$. Dan levert dit in het totaal van alle $72 + 69 = 141$ deelnemende scholen nog steeds een veel hoger gemiddelde op dan $5\frac{1}{2}$. Met andere woorden, de rekenvaardigheid in de deelnemerspopulatie is veel hoger dan de rekenvaardigheid in de totale populatie van alle scholen in Nederland.

Het bovenstaande laat zien dat het onverantwoord is om TIMSS-scores van een land met een te geringe respons op de oorspronkelijke steekproef geldig te verklaren voor de gehele populatie van het desbetreffende land. Er bestaat immers de mogelijkheid dat er een correlatie is tussen het weigeren van geselecteerde scholen om aan TIMSS mee te doen en de rekenvaardigheid op die scholen.

Alle redenen tot zorg

Om twee redenen telt de mogelijke correlatie tussen weigeren en rekenvaardigheid extra zwaar voor de Nederlandse deelname. In de eerste plaats was de respons in Nederland *uitzonderlijk laag*: slechts 48 procent, terwijl de andere zes landen die onder de vereiste 85 procent zaten, percentages hadden van 70 (USA), 71 (Denemarken), 77 (Schotland), 81 (Marokko en Hong Kong) en 83 (Engeland)

(p. 180 in [4]). Overigens, ook in TIMSS 2003 voldeed Nederland niet aan de representativiteitseis. Van de 150 scholen in de oorspronkelijke steekproef deden er toen maar 77 aan het onderzoek mee (52 procent). Er zijn toen 53 reservescholen toegevoegd. In 2003 waren er 26 deelnemende landen in *grade 4*. Zeven landen, waaronder Nederland zaten toen onder 85-procentnorm. De andere waren Australië (78%), Filippijnen (78%), Hong Kong (77%), USA (70%), Schotland (64%) en Engeland (54%). Ook toen had Nederland dus het laagste deelnamepercentage in de oorspronkelijke steekproef (pp. 355 en 359 in [2]).

De tweede reden is de *uitzonderlijk lage waardering* voor rekenen en wiskunde in de huidige vorm onder Nederlandse scholieren en leerkrachten. Iedereen die de rekendiscussie in de media volgt, weet hoe groot dit probleem bij ons is, maar de Nederlandse TIMSS-resultaten onderstrepen dit ook weer. Bovendien zijn er bij ons grote verschillen tussen scholen. In een recent rapport schat de onderwijsinspectie het aantal rekenzwakke scholen op ongeveer een kwart.

Al met al lijkt het geen onredelijke veronderstelling dat rekenzwakke scholen eerder geneigd zijn niet aan TIMSS mee te doen dan rekensterke scholen. Dit zou betekenen dat de TIMSS-resultaten van Nederland geflatteerd zijn. Maar in elk geval is het onverantwoord om, zoals voorstanders van zogenaamd 'realistisch' rekenonderwijs doen, uit de TIMSS-scores te concluderen dat Nederlandse kinderen internationaal gezien uitstekende rekenaars zijn en dat het rekenonderwijs op de Nederlandse basisscholen van hoge kwaliteit is. De tegengestelde conclusie, namelijk dat er alle redenen tot zorg is over ons rekenonderwijs, lijkt eerder door de TIMSS-resultaten te worden ondersteund. Denk daarbij vooral ook aan de geringe waardering voor rekenen en aan de onrustbarende verschillen tussen jongens en meisjes en tussen allochtone en autochtone leerlingen in de Nederlandse deelnemersgroep.

Referenties

- M.R.M. Meelissen en M. Drent, 2008, *TIMSS-2007 Nederland*, Universiteit Twente, Vakgroep Organisatie en Management, <http://doc.utwente.nl/60205/1/TIMSS-2007.pdf>
- Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004), *TIMSS 2003 International Mathematics Report. Findings from IEAs Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. <http://timssandpirls.bc.edu/timss2003/mathD.html>
- Mullis, I.V.S., Martin, M.O., & Foy, P. (2008). *TIMSS 2007 International Mathematics Report. Findings from IEAs Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston: Boston College, TIMSS & PIRLS International Study Center, <http://timss.bc.edu/TIMSS2007/mathreport.html>
- Olson, J.F., Martin, M.O., & Mullis, I.V.S. (Eds.). (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College, <http://timss.bc.edu/TIMSS2007/techreport.html>
- Rekenpeil basisonderwijs langzaam achteruit*, gezamenlijk persbericht NWO en Universiteit Twente, 9 december 2008, <http://www.nwo.nl/nwohome.nsf/pages/NWOA.7M6GUP>