



# De onmiskenbare daling van het prestatiepeil bij de bewerkingen sinds 1987

- een reactie -

C.M. van Putten  
Universiteit Leiden

*Treffers' bewering in zijn artikel 'De kwaliteit van het reken-wiskundeonderwijs' (2007) dat de periodieke rekenpeilingen voor de bewerkingen niet valide zijn en er dus geen sprake is van kwaliteitsachteruitgang in het rekenonderwijs, is onhoudbaar. Want uit het feit dat individuele toetsafnames betere resultaten opleverden, kan immers niet geconcludeerd worden dat er geen sprake is van achteruitgang van het rekenpeil. De aanname dat de PPON-opgaven in volgorde van moeilijkheidsgraad werden aangeboden en daardoor het hoofdrekenen ook bij moeilijkere opgaven uitlokte, waardoor prestaties verminderden, klopt niet. Bovendien gaat Treffers' betoog voorbij aan enkele relevante bevindingen. Zo bleken niet alleen sterke maar ook zwakke en gemiddelde leerlingen beter te staartdelen dan happend te rekenen. En het nieuwe-methode-effect was over het algemeen klein en was juist bij de bewerkingen in de peilingen van 2004 afwezig tot negatief ten opzichte van eerdere peilingen. Ook voor elke afzonderlijke rekenstrategie bleken de prestaties op de deelsommen tussen 1997 en 2004 achteruit te zijn gegaan.*

*Door deze omissies en door de standaarden voor rekenprestaties af te wijzen, draagt Treffers niet bij aan oplossingen voor de serieuze problemen in het Nederlandse rekenonderwijs. Deze problemen worden opnieuw duidelijk uit een nadere analyse van de bewerkingsopgaven voor vermenigvuldigen uit PPON. Alleen sterke rekenaars bleken in 2004 met alle strategieën succesvol op de opgave  $99 \times 99$ , maar gemiddelde en zwakke rekenaars boekten alleen succes met de traditionele vermenigvuldiging.*

## 1 Inleiding

Treffers (2007) noemt 'de hele commotie over de teruggang van het cijferen ... een storm in een glas water' (pag.14). Hij ziet geen reden om te veronderstellen dat de opbrengst van het schriftelijk rekenen in 2004 ten opzichte van 1987 is teruggelopen, omdat de peilingen voor dit onderwerp 'niet valide' (pag.13) zouden zijn. Het gaat hier om de Periodieke Peiling van het Onderwijsniveau (PPON) bij rekenen op het einde van de basisschool (Janssen, Van der Schoot & Hemker, 1999 en 2005). Treffers cijfert de sterke daling van de prestaties bij de bewerkingen in de periode 1987-2004 volledig weg. Ook hecht hij er weinig belang aan dat de peilingen achterblijven bij de standaarden die zijn afgeleid uit de ministeriële eindtermen. In deze reactie bespreek ik de feitelijke grondslag van zijn betoog en probeer ik aanmerkelijk te maken dat er wel degelijk sprake is van serieuze problemen in het Nederlandse rekenonderwijs.

## 2 Standaarden voor rekenen

Treffers haalt auteurs aan die de standaarden uit de PPON-

rapporten sterk relativeren: de lat zou veel te hoog gelegd worden en het zou om te subjectieve oordelen gaan die irreële verwachtingen wekken. Het gaat om streefdoelen die we niet serieus hoeven te nemen en daarmee is volgens Treffers de kous af.

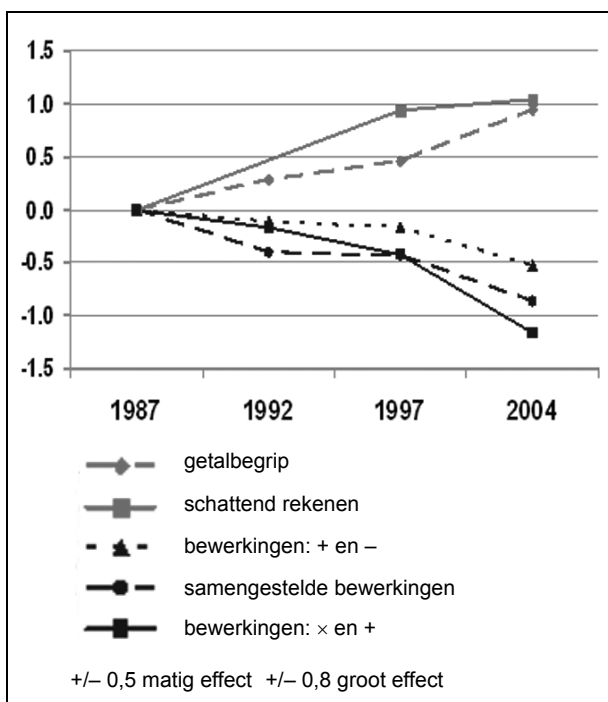
Deze afwijzing van de PPON-standaarden als beoordelingskader voor het Nederlandse rekenonderwijs, doet echter geen recht aan de zorgvuldige wijze waarop het Cito getracht heeft een kader te bepalen om de gepeilde prestaties tegen af te zetten (Van der Schoot, 2008). Twee panels van 25 beoordelaars (merendeels ervaren leerkrachten uit groep 8) hebben in een aantal rondes uitspraken gedaan over de minimale, voldoende en gevorderde beheersing van twintig tot dertig opgaven. Dit oordeel hebben zij moeten herhalen na discussie met andere beoordelaars en (opnieuw) na kennisname van de feitelijke beheersing van elke opgave in de leerlingpopulatie. Natuurlijk valt op elke kaderstelling wel wat af te dingen. Maar eenieder die zich afvraagt of het behaalde prestatieniveau aan de officieel gestelde doelen voldoet, kan niet om de conclusie heen dat de rekenprestaties over het algemeen tegenvallen. Te veel kinderen zijn op het einde van het basisonderwijs niet in staat om heel gewone rekenopgaven goed te beantwoorden. Een voorbeeld: de gemiddelde leerling uit PPON 2004 beheerste de volgende opgave (die met papierhulp mocht worden beantwoord) niet: Hoeveel kosten 25 atlanten van € 19,50 per stuk?

Naar mijn mening bagatelliseert Treffers hier een serieus probleem, dat echter door velen als zodanig onderkend wordt.

### 3 Veranderingen in het peil van de Nederlandse rekenprestaties

Zou het peil van het Nederlandse rekenonderwijs in de afgelopen twintig jaar sterk zijn gestegen, dan was dit achterblijven bij de standaarden misschien minder erg: het peil zou dan ten minste verbeterd zijn. Maar dat is niet het geval. Op de meeste onderdelen van het rekenen zijn er slechts kleine veranderingen in het peil, soms ietsje vooruit en soms ietsje achteruit. Op maar twee onderdelen is er sprake van een zeer groot positief effect: bij getalbegrip en bij schattend rekenen (fig.1).

Toch bereikte in 2004 maar 42 procent van de leerlingen de standaard voldoende bij deze onderwerpen (de standaard voldoende houdt in dat 70 à 75 procent van de leerlingen een bepaalde peilscore zou moeten overtreffen). Treffers beweert overigens dat er ook bij het hoofdrekenen als geheel een ‘aanzienlijk’ (pag.12) positief effect optrad. Maar alleen bij het hoofdrekenend optellen en aftrekken trad een matig grote verbetering op tot 1997; het peil van hoofdrekenend vermenigvuldigen en delen daalde juist iets.



figuur 1: verandering rekenpeil 1987-2004  
Cito periodieke peilingen (PPON)

Bij het rekenen met procenten trad nog een matig grote verbetering op tot 1997, terwijl de kleine verbeteringen bij verhoudingen en breuken tot 1997 daarna juist weer lijken terug te lopen. Bij procenten bereikte 58 procent van de leerlingen in 2004 de standaard voldoende; dat was in 1997 nog ongeveer 40 procent. Ook in dit opzicht een verbetering dus.

Daarentegen is er bij alle onderdelen van de bewerkingen (voorheen cijferen) een duidelijke teruggang te zien in de periode 1987-2004 (fig. 1): bij optellen en aftrekken een matige teruggang, bij samengestelde bewerkingen een grote teruggang en bij vermenigvuldigen en delen een zeer grote teruggang. Bij deze onderdelen behaalden in 2004 respectievelijk slechts 27, 16 en 12 procent de standaard voldoende. Dit was in 1997 ongeveer 50 procent (optellen en aftrekken), ongeveer 55 procent (samengestelde bewerkingen) en iets meer dan 50 procent (vermenigvuldigen en delen); (Janssen e.a., 1999 en 2005).

### 4 Aanvullend onderzoek naar strategiegebruik in PPON

Het is deze sterke achteruitgang bij de bewerkingen die door Treffers volledig wordt weggecijferd. Hij gebruikt hiervoor de Leidse onderzoeksresultaten van Van Putten (2005) en Van Putten en Hickendorff (2006), die met de PPON-boekjes voor bewerkingen uit 1997 en 2004 aanvullende analyses hebben uitgevoerd. Alle toetsboekjes zijn bekeken op uitwerkingen van leerlingen bij hun antwoorden en daar kwam het volgende uit.

De achteruitgang van de prestaties bij de deelsommen tussen 1997 en 2004 zou deels verklaard kunnen worden door het veranderde strategiegebruik van leerlingen: de relatief succesvolle staartdeling werd in 2004 veel minder vaak toegepast dan in 1997, terwijl het gebruik van de iets minder succesvolle realistische strategieën vrijwel niet toenam.

Wat echter wel sterk is toegenomen, is het verschijnsel dat leerlingen opgaven beantwoordden zonder een uitwerking op te schrijven, en meestal was dat antwoord dan fout (in 2004 nog vaker dan in 1997). Over deze toetsboekjes beweerde Treffers overigens in Trouw (19 december 2007) dat er ‘geen ruimte (is) voor bewerkingen, dus daarom schrijven ze niets op’. Dat is echter niet het geval. Er is voldoende ruimte in de PPON-toetsboekjes en er zijn dan ook meer dan 2100 staartdelingen, happenschema’s en andere uitwerkingen aangetroffen bij de antwoorden op de 5704 delingsopgaven van de ruim 1000 leerlingen in 2004 (Van Putten, 2005, pag.127).

## 5 Groepsgewijze en individuele toetsafname in PPON

Uit een analyse op twee PPON-deelopgaven die ook individueel waren afgenomen bij een klein deel van de leerlingen, bleken de prestaties met een testleider ernaast veel beter dan die onder de gewone groepsgewijze PPON-omstandigheden. Antwoorden zonder uitwerking kwamen met een testleider erbij vrijwel niet meer voor (Van Putten & Hickendorff, 2006).

Door simpelweg twee opgaven uit (de gewone) PPON 1987 te nemen, volgens Treffers met overeenkomstige moeilijkheidsgraad, en deze qua goedpercentage te vergelijken met de beide individueel afgenomen opgaven uit 2004, concludeert hij dat de achteruitgang in prestaties een storm in een glas water is. De PPON-peiling voor de bewerkingen is volgens hem niet valide, want ze geeft niet weer wat leerlingen in feite wel kunnen als ze gedwongen zouden worden om ook een schriftelijke berekening te geven (pag.13-14).

Naar mijn mening is deze redenering van Treffers onhoudbaar. Hij gaat uit van veronderstellingen die aantoonbaar niet kloppen en hij zet bovendien de zaak op zijn kop. Om met het laatste te beginnen. Het realistische rekenonderwijs heeft juist als kernpunt de diversiteit aan rekenstrategieën van leerlingen. Die diversiteit zien we dan ook terug in de peilingen: traditionele naast realistische oplossingsstrategieën, volledig uitgeschreven berekeningen naast kleine geheugensteuntjes, en zoals gezegd, ook vele antwoorden zonder enige uitwerking erbij. Zo rekenden de leerlingen anno 2004 en de PPON-resultaten maken daar de balans van op. Te stellen, zoals Treffers doet, dat in PPON de leerlingen geïnstrueerd zouden moeten worden om bij de bewerkingen alleen maar schriftelijk te rekenen, zou peilingen opleveren die niet weerspiegelen hoe leerlingen van nu hebben leren rekenen. Dat zou pas een invalide meting opleveren.

## 6 Veronderstellingen en omissies

Ik mis in Treffers' betoog een kritische instelling, gericht op het precies willen nagaan wat er aan de hand is met het hedendaagse rekenonderwijs in de Nederlandse basisscholen, zeker als bepaalde resultaten verslechteren en leerlingen in toenemende mate strategieën vertonen die riskant zijn. Maar Treffers' betoog is gebouwd op veronderstellingen die onjuist zijn. Ook worden gepubliceerde onderzoeksresultaten genegeerd die kennelijk niet uitkomen. Hier volgen de belangrijkste problemen in zijn betoog.

- Bij de twee individueel afgenomen peilingsopgaven

voor delen was de goedscore 30 procentpunten hoger dan die in de gewone peilingstoets. Met die 30 procent gaat Treffers aan het rekenen om de achteruitgang van het peil in PPON ongedaan te maken. Maar hier moet in ogenschouw worden genomen dat er sprake was van een bijzondere individuele afname. Het doel was te achterhalen welke strategie de leerling hanteerde en de testleider vroeg door als dat onduidelijk was. De uitkomsten van een dergelijke afname geven daarom vooral informatie over het strategiegebruik, maar zeggen minder over het prestatiepeil. De algemene testliteratuur noemt enkele risico's van groepstests die tot lagere scores kunnen leiden, zoals motivatieverlies en het niet volgen van de instructies (Gregory, 2004). Dergelijke problemen worden tijdens individuele afnamen meestal voorkomen door de testleider. Dat zien we ook bij PPON gebeuren: het uitvallen van opgaven door overslaan, door verkeerd gekozen bewerkingen en door te onduidelijke uitwerkingen daalde van 19 procent in de groepsgewijze afname tot vrijwel nul in deze individuele afnames. Alleen al hierdoor steeg de prestatie in deze laatste conditie. Bovendien betrof dit individuele onderzoek maar twee delingsopgaven op een beperkt aantal leerlingen (140). Allemaal redenen om het prestatieaspect van deze gegevens met grote voorzichtigheid te behandelen. Dat doet Treffers echter niet.

- Waarom geven leerlingen tegenwoordig zo vaak een antwoord zonder een uitwerking of berekening? Treffers beaamt dat de grotere aandacht voor het hoofdrekenen een rol kan hebben gespeeld, maar er is volgens hem nog iets anders in het spel. Hij veronderstelt dat de peilingsopgaven beginnen met gemakkelijke sommen, die goed uit het hoofd te doen zijn, en dat leerlingen daarna bij de moeilijkere sommen in het hoofdrekenspoor blijven hangen. Hij geeft voorbeelden van opgaven. Maar Treffers beseft niet dat de volgorde van de opgaven in peilingstoetsen heel anders is dan in de PPON-rapportage. In PPON-2004 zijn er acht verschillende toetsboekjes met daarin verschillende subsets van deelopgaven in verschillende volgordes.

In de PPON-rapporten zijn de opgaven per onderdeel voor de overzichtelijkheid geordend op de peilstok naar moeilijkheidsgraad. Maar de volgorde in de toetsen is beslist geen ordening naar moeilijkheidsgraad. In het toetsboekje 'Bewerkingen: vermenigvuldigen en delen' uit PPON-1997 zit de lastigste deelopgave zelfs in het begin van de toets.

Er waren trouwens in beide peilingsjaren meer leerlingen die hoofdrekenen bij de ene opgave combineerden met een geschreven strategie bij een andere opgave, dan leerlingen die alleen maar hoofdrekenden (Janssen e.a., 2005, pag.130). Daaraan is al te zien dat het vastkleven aan hoofdrekenen maar betrekkelijk is. Treffers' onjuiste veronderstelling verklaart bovendien

niet waarom jongens vaker uit het hoofd rekenden dan meisjes en zwakke rekenaars vaker dan gemiddelde en sterke rekenaars (Janssen e.a., 2005, pag.127/8; Van Putten & Hickendorff, 2006, pag.21).

- Treffers stelt dat staartdelingen alleen maar betere resultaten te zien geven dan realistische uitwerkingen, omdat de betere leerlingen vaker staartdeelden dan de minder goede rekenaars (pag.14). Hij gaat dan voorbij aan de gepubliceerde figuren van de Leidse onderzoekers (Janssen e.a., 2005, pag.128/9) waarin de resultaten zijn uitgesplitst naar het algemeen rekenniveau van de leerlingen. Daaruit blijkt dat ook binnen elke niveaugroep (zwak, gemiddeld, sterk) de staartdeling het iets beter deed, zowel in 1997 als in 2004. De vaak geuite verwachting (bijvoorbeeld door M. Kool tijdens de 'Nationale Rekentest' van december 2007) dat het realistische happenschema juist voor zwakke rekenaars tot zoveel betere resultaten leidt dan de traditionele staartdeling, wordt op geen enkele wijze ondersteund door onze onderzoeksresultaten. Op het einde van deze reactie zullen we ook bij cijferend vermenigvuldigen zien dat voor zwakke rekenaars de traditionele vermenigvuldiging tot betere resultaten leidde dan de realistische strategieën.
- Treffers beweert dat het gerapporteerde positieve effect van nieuwe reken-wiskundemethoden in strijd is met de prestatieergang bij de bewerkingen. Dat is echter niet het geval en daarom is er ook geen raadsel dat moet worden opgelost. Jaareffecten (verandering van rekenpeil tussen bijvoorbeeld PPON 1997 en 2004) en methode-effecten (verschillen in peil tussen leerlingen die met andere methoden les kregen) blijken onafhankelijk van elkaar te zijn (Janssen e.a., 1999, pag.179 en 187) en ze kunnen bovendien tegelijkertijd optreden. De jaareffecten die het Cito rapporteerde in de peilingen van 1992 en 1997 zijn gecorrigeerd voor verschuivingen in methodegebruik (niet in de peiling van 2004 omdat toen ongeveer 80 procent van de scholen vanwege de invoering van de euro een nieuwe rekenmethode had aangeschaft). In tegenstelling tot wat Treffers beweert (pag.12), is dit nieuwe-methode-effect in 2004 juist bij de bewerkingen vermenigvuldigen en delen afwezig ten opzichte van 1997 en is het negatief ten opzichte van 1992 (Janssen e.a., 2005, pag.237). Bovendien zijn de overige positieve nieuwe-methode-effecten over het algemeen klein (pag.236).
- Treffers beperkt zich bij het bespreken van de neergang van de bewerkingsprestaties tot de door ons geconstateerde veranderingen in strategiegebruik. Maar hij negeert daarbij een ander gerapporteerd onderzoeksresultaat uit Leiden: ongeacht het strategiegebruik gingen de prestaties tussen 1997 en 2004 naar beneden. Elke onderscheiden strategie (staartdeling, happenschema, hoofdrekenden) werd in de vierde pei-

ling slechter uitgevoerd dan in de derde peiling (Janssen e.a., 2005, pag.128/9). Ook dit is een belangrijk (negatief) resultaat dat om een nadere verklaring vraagt. Zo hebben velen zich de vraag gesteld of het didactische evenwicht tussen inzicht nastreven en vaardigheden laten verwerven niet te veel naar het eerste is verschoven, ten koste van het tweede.

---

## 7 Hoe krijgen we de geest weer op het papier?

Ik ben resumerend van mening dat Treffers' bagatellisering van de prestatieergang bij cijferend vermenigvuldigen en delen, optellen en aftrekken niet in overeenstemming is met de feiten. Er is sprake van serieuze problemen in een centraal deel van de rekenvaardigheid op het einde van de basisschool. Dit vraagt om nader onderzoek van en bezinning op de huidige lespraktijken en reken-wiskundemethoden, en om verdere verbetering van de kwaliteit van het rekenonderwijs. Het zal in de praktijk van het rekenonderwijs nog een hele krachtsinspanning vergen om de geest die vermoedelijk door het vele hoofdrekenden uit de fles is ontsnapt, weer op het papier te krijgen bij de 'bewerkingen', zeker bij jongens en bij zwakke rekenaars. Het goed leren noteren van berekeningen en oplossingen is een van de onderwerpen die K. Buys heeft onderzocht voor zijn proefschrift, waarop hij op 19 mei 2008 hoopt te promoveren. Welke oplossingsstrategieën leerlingen gebruiken wanneer zij rekenopgaven beantwoorden zonder enige uitwerking, is recent onderzocht door Hickendorff e.a. (2008).

---

## 8 Strategiegebruik en prestatie bij vermenigvuldigen in PPON

Wij zijn in Leiden inmiddels weer een stap verder met aanvullende analyses op het PPON-materiaal. Na de deelopgaven zijn nu ook de vermenigvuldigopgaven uit PPON 1997 en 2004 geanalyseerd (en dit voorjaar worden de uitwerkingen bij de aftrekopgaven uit PPON 2004 gecategoriseerd). In grote lijnen volgen de resultaten bij bewerkingen-vermenigvuldigen die bij het delen: een sterke afname van de prestaties, samenhangend met een afname van het gebruik van traditionele procedures en een toename van het hoofdrekenden (Van Putten & Hickendorff, 2007).

Figuur 2 toont de percentages goede antwoorden voor de opgave ( $99 \times 99 =$ ) uitgesplitst naar gebruikte strategie en rekenniveau van de leerlingen. 31 Procent van de leerlingen gebruikte het traditionele vermenigvuldigalgoritme, 19 procent een realistische strategie (meestal com-

Gebruikte strategie	Algemeen rekenniveau			Alle leerlingen
	zwak	midden	sterk	
traditioneel algoritme	45%	71%	90%	71%
realistische strategie	5%	27%	58%	35%
zonder uitwerking	14%	27%	61%	34%
alle strategieën	20%	41%	67%	43%

figuur 2: percentage goede antwoorden op de opgave  $99 \times 99 =$  naar gebruikte strategie en algemeen rekenniveau ( $N = 371$ )

penseren), 39 procent gaf een antwoord zonder uitwerking en 10 procent sloeg de opgave over.

Er is speciaal ingezoomd op deze opgave omdat hij zich zo goed lijkt te lenen voor een realistische aanpak. Zolang de leerlingen maar traditioneel rekenden, ging het meestal goed (in 71 procent van de gevallen). Realistische aanpakken via bijvoorbeeld  $100 \times 99$  of  $100 \times 100$  leidden in het algemeen tot foutieve uitwerkingen en antwoorden.

Het begon met problemen in  $100 \times 99$  of  $100 \times 100$  (veel fouten als: 990, 1000 of 1.000.000) en vervolgens trad het probleem op hoeveel daarvan af te trekken (compenseren, met veel fouten als:  $-100$ ,  $-198$ ,  $-18$ ,  $-2$  of  $-1$ ). Dit soort fouten was ook in de antwoorden van de hoofdrekkenaars terug te vinden. Eigenlijk was alleen de traditionele aanpak hier succesvol en konden slechts de sterke rekkenaars (de beste 33 procent) zich een realistische of hoofdrekkenaanpak veroorloven. Alle andere combinaties (omcirkeld) waren kansloos.

Deze bewerkingsopgave bleek ook lastig op de 'Nationale Rekentest' van december 2007, omdat hij daar onder tijdsdruk uit het hoofd gedaan moest worden. Twee van de drie finalisten bezweken en toen kreeg de derde de gelegenheid om het goed uit te rekenen:  $(100 \times 99) - 99 = 9801$ .

## 9 Getalbegrip, inzicht en rekenvaardigheid?

Het bovenstaande roept de vraag op waarom getalbegrip, inzicht en schattend rekenen zo weinig leerlingen op het juiste pad hielden. Deze speerpunten van het realistisch rekenen laten op de peilingen grote vooruitgang zien, maar lijken bij deze opgave niet goed gefunctioneerd te hebben. Waarom ging  $100 \times 100$  of  $100 \times 99$  zo vaak fout? Waar was het getalbegrip? En waarom werd er zo vaak een onjuiste compensatie gebruikt? In het opper-

vlaktemodel van  $100 \times 100$  tegels verwijder je eerst 100 en daarna 99 tegels, zodat je op 9801 uitkomt voor een oppervlakte van  $99 \times 99$  tegels. In het PPON-materiaal heb ik geen enkele leerling een dergelijke visualisatie zien gebruiken. Waarom fuctioneerde het inzichtelijk rekenen hier niet?

Inzichtelijk en handig rekenen bestaat uiteindelijk uit een grote verzameling van getalkennis en uit een diversiteit aan alternatieve rekenvaardigheden. Compenseren is daar een voorbeeld van, maar veel leerlingen blijken er grote moeite mee te hebben. Als getalfeiten niet gekend en vaardigheden onvoldoende verworven worden, dan gaat het mis, niet alleen bij vermenigvuldigen maar ook bij andere bewerkingen.

### Literatuur

- Gregory, R.J. (2004). *Psychological testing: history, principles, and applications*. Fourth edition. Boston, MA: Pearson Education Group, Inc.
- Hickendorff, H., W.J. Heiser, C.M. van Putten & N.D. Verhelst (2008). *Written and mental computation on division problems: Strategy choice and accuracy in a partial Choice/No-Choice design*. (Submitted for publication).
- Janssen, J., F. van der Schoot, B. Hemker & N. Verhelst (1999). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 1997. Periodieke Peiling van het Onderwijsniveau, PPON-reeks nummer 13*. Arnhem: Cito.  
[http://www.cito.nl/share/PPON/Cito\\_pponbalans\\_13.pdf](http://www.cito.nl/share/PPON/Cito_pponbalans_13.pdf)
- Janssen, J., F. van der Schoot & B. Hemker (2005). *Balans (32) van het reken-wiskundeonderwijs aan het einde van de basisschool 4. Uitkomsten van de vierde peiling in 2004. Periodieke Peiling van het Onderwijsniveau*. Arnhem: Cito.  
[http://www.cito.nl/share/PPON/Cito\\_pponbalans\\_32.pdf](http://www.cito.nl/share/PPON/Cito_pponbalans_32.pdf)
- Putten, C.M. van (2005). Strategiegebruik bij het oplossen van deelsommen. In: J. Janssen, F. van der Schoot & B. Hemker. *Balans (32) van het reken-wiskundeonderwijs aan het einde van de basisschool 4. Periodieke Peiling van het Onderwijsniveau*. Arnhem: Cito (125-131).
- Putten, C.M. van & M. Hickendorff (2006). Strategieën van leerlingen bij het beantwoorden van deelopgaven in de periodieke peilingen aan het eind van de basisschool van 2004

- en 1997. *Reken-wiskundeonderwijs: onderzoek, ontwikkeling, praktijk*, 25(2), 16-25.
- Putten, C.M. van & M. Hickendorff (2007). *Strategiegebruik en prestaties bij vermenigvuldigen in groep 8 in PPON 1997 en 2004*. Presentatie op de Onderwijs Research Dagen, Groningen (6 juni 2007).
- Schoot, F. van der (2008). *Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPON*. Arnhem: Cito. [http://www.cito.nl/po/ppon/alg/Cito\\_PPON\\_20\\_jaar.pdf](http://www.cito.nl/po/ppon/alg/Cito_PPON_20_jaar.pdf)
- Treffers, A. (2007). De kwaliteit van het rekenonderwijs; een virtueel vraaggesprek. *Reken-wiskundeonderwijs: onderzoek, ontwikkeling, praktijk*, 26(4), 11-17.

---

*Recently Treffers (2007) stated that the periodical national mathematics assessments (PPON) were invalid for the domain of written or multicolumn division, multiplication, addition, and subtraction. So he saw no reason to conclude that achievement had dropped considerably in this respect. My reaction is that this position is not defensible, because a number of his interpretations of the PPON results are incorrect or biased. Inferring a good achievement level from individual testing research into pupils' strategy use ignores important differences in testing conditions compared to standard group assessment in PPON. Treffers assumed that PPON tests presented easy items before difficult ones, thus stimulating mental calculation even with difficult items. However, this was not the case: item order was mixed according to difficulty level. Treffers neglected the published result that weak calculating pupils also achieved better when using the traditional long division algorithm compared to the realistic 'take away' strategy. Regarding a general positive new textbook effect in the assessments of 2004, it was not mentioned that for the division domain this effect did not actually exist compared to PPON 1997, and that this effect was even negative compared to PPON 1992. Nor did he mention that the achievement drop for written division also occurred for each separate type of solution strategy (traditional, realistic, mental). Treffers' negative attitude towards achievement standards has led him to deny serious problems in the heart of Dutch mathematics education. New analyses on PPON assessments for written multiplication show again the seriousness of these problems. Only strong calculating pupils were successful on the PPON 2004 item  $99 \times 99$  irrespective of their solution strategy. Only the traditional written multiplication brought success for average and weak calculating pupils; all other combinations of strategy use and mathematics level gave disastrous results.*