

Peilingsonderzoek rekenen-wiskunde in het basisonderwijs – Balans van 25 jaar PPON

De periodieke peiling van het onderwijsniveau in Nederland (PPON) bestaat in 2016 ruim 25 jaar. Een van de doelen van peilingsonderzoek is het leveren van een empirische basis voor de maatschappelijke discussie over inhoud, kwaliteit en niveau van het onderwijs, in dit geval op de basisschool.

In deze bijdrage willen we het belang van het ontwerp van PPON benadrukken en gaan we in op de lessen die we kunnen leren op basis van die uitkomsten voor de inrichting en het ontwerp van toekomstig peilingsonderzoek bij rekenen-wiskunde. We vergelijken daarbij het ontwerp van PPON met dat van andere peilingsonderzoeken, zoals TIMSS en PISA. Tot slot staan we stil bij de toekomst van PPON.

Als we de resultaten van de rekenpeilingen van PPON over de afgelopen 25 jaar overzien kunnen we de volgende conclusies trekken:

- Er is in de peilingen van PPON een forse dynamiek waargenomen in de ontwikkeling van de rekenvaardigheid.
- Om die dynamiek te kunnen waarnemen is het cruciaal welke rapportage-eenheden worden gekozen. Of het niveau stijgt dan wel daalt, kan niet eenduidig geconcludeerd worden.
- Samenvoegen van rapportage-eenheden verdoezelt de eventueel aanwezige effecten.

Juist omdat er bij PPON in de huidige opzet een fijnmazige meting wordt verricht is het mogelijk om de veranderingen in rekenvaardigheid op te merken. Dit is onmisbaar om een gedegen discussie te voeren over het niveau van het rekenonderwijs in Nederland.

INLEIDING

De periodieke peiling van het onderwijsniveau in Nederland bestaat in 2016 ruim 25 jaar en er zijn in die periode vijf cycli uitgevoerd voor rekenen-wiskunde, zowel in jaargroep 8 als in jaargroep 5. Ook zijn er inmiddels vier periodieke peilingen uitgevoerd in de eindgroep van de speciale basisschool. Vanaf 2014 is de regie van het peilingsonderzoek in het basisonderwijs in handen gegeven van de Inspectie van het Onderwijs (Onderwijsinspectie). Dat is aanleiding voor een terugblik.

Eén van de doelen van peilingsonderzoek is het leveren van een empirische basis voor de maatschappelijke discussie over inhoud, kwaliteit en niveau van het onderwijs, in dit geval op de basisschool. De uitkomsten van de peilingen bij rekenen-wiskunde hebben zeker aan dat doel beantwoord. Ze hebben geleid tot stevige discussies over het niveau van de rekenprestaties in Nederland en ze speelden een belangrijke rol bij de tot stand koming van de referentieniveaus voor rekenen.

In deze bijdrage willen we belang van het ontwerp van PPON benadrukken en gaan we in op de lessen die we kunnen leren op basis van die uitkomsten voor de inrichting en het ontwerp van toekomstig peilingsonderzoek bij rekenen-wiskunde. We vergelijken daarbij het ontwerp van PPON met dat van andere peilingsonderzoeken, zoals TIMSS en PISA. Dat is relevant nu de regie van peilingsonderzoek in Nederland in handen is gekomen van de inspectie van het onderwijs.

Het gaat in dit artikel dus niet zozeer om de vraag wat die peilingen voor het rekenonderwijs hebben opgeleverd en het is ook geen inhoudelijke bespreking van de uitkomsten van de

Jan van Weerden,
Jan Janssen & Floor
Scheltens
[Cito, Arnhem](#)

Van Weerden,
J., Janssen, J., &
Scheltens, F. (2016).
[Peilingsonderzoek
rekenen-wiskunde in
het basisonderwijs
– Balans van 25 jaar
PPON. Volgens Bartjens
– ontwikkeling en
onderzoek, 35\(4\), 41-50.](#)

opeenvolgende peilingen, dat is reeds door anderen in diverse publicaties gedaan (bijvoorbeeld KNAW, 2009; Van Putten & Hickendorff, 2006). Tot slot staan we stil bij de toekomst van PPON en geven we enkele ervaringen weer uit de workshop die over deze materie is gehouden op de Panamaconferentie van januari 2015.

PEILINGSONDERZOEK EN DIDACTISCH BETEKENISVOLLE EENHEDEN

Peilingsonderzoek is steekproefonderzoek en bestaat naast een inventarisatie van het onderwijsaanbod bij de leerkrachten uit een uitgebreide verzameling opgaven waarmee de onderwijsprestaties van de leerlingen in kaart kan worden gebracht. Bovendien worden er een aantal achtergrondvariabelen van leerlingen meegenomen, zodat gerapporteerd kan worden over verschillen tussen groepen (bijvoorbeeld op basis van geslacht, formatiegewicht, leertijd). Van meet af aan is benadrukt dat in het PPON-ontwerp een ruime dekking van de leer- en vormingsgebieden voorop moest staan om op basis van de resultaten een steekhoudende inhoudelijke discussie over het onderwijs te kunnen voeren (Van der Schoot, 2008; Van Weerden & Hiddink, 2013). PPON streeft er naar de vaardigheden van de leerlingen te beschrijven op basis van compacte, didactisch betekenisvolle eenheden (DBE-en), oftewel 'indivisible curricular elements' (Bock, Mislevy & Woodson, 1982), hetgeen betekent dat er niet wordt volstaan met algemene uitspraken over bijvoorbeeld het niveau van het reken-wiskundeonderwijs, maar dat er gerapporteerd wordt over kleinere, herkenbare onderwerpen binnen dat onderwijs.

Deze didactische betekenisvolle eenheden vormen meetschalen, waarmee de vaardigheid van groepen leerlingen gedetailleerd in kaart kunnen worden gebracht. Zo is bij een vergelijking van de leerprestaties over verloop van tijd duidelijk te maken bij welke onderwerpen er vooruitgang, achteruitgang, dan wel stabilisatie optreedt. Hoe meer onderwerpen er betrouwbaar en valide gemeten worden en afzonderlijk kunnen worden gerapporteerd, hoe meer er in gedetailleerde vorm iets gezegd kan worden over de kwaliteit van het onderwijs.

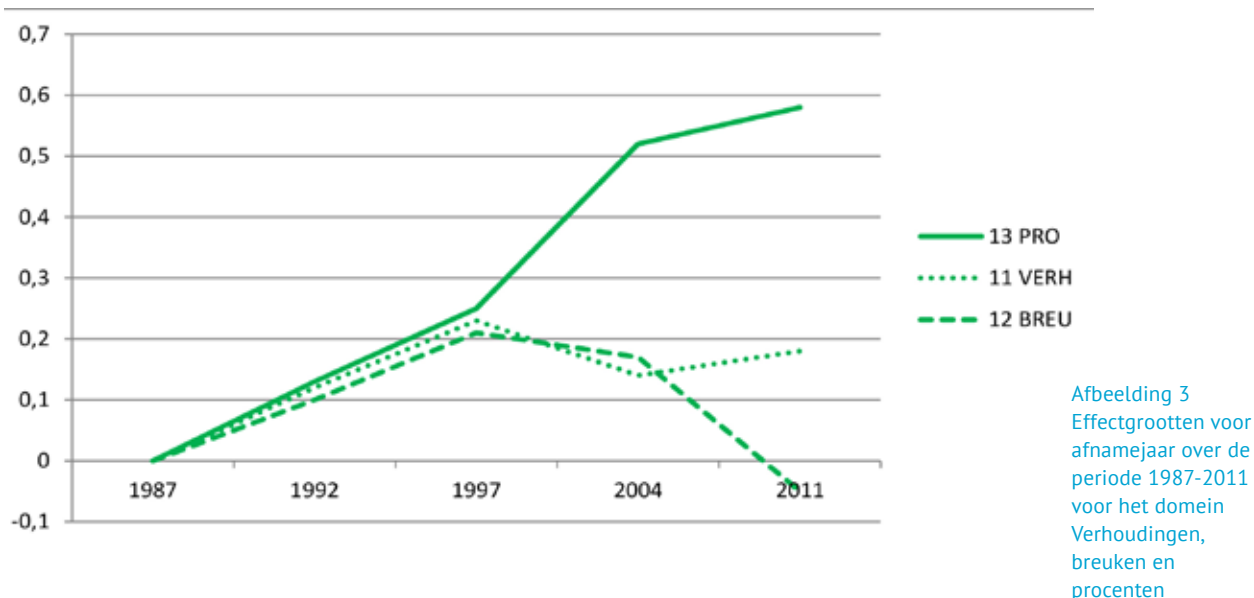
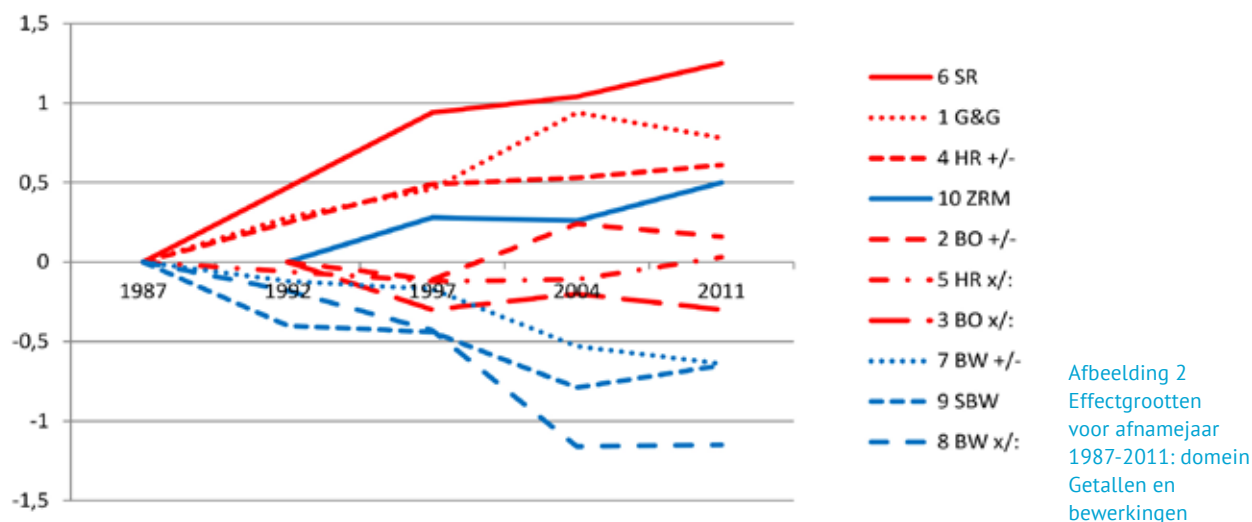
Dit valt eenvoudig te illustreren met de uitkomsten van de vijf reken-wiskundepeilingen in jaargroep 8 uitgevoerd in de afgelopen 26 jaar, waarbij er gerapporteerd kan worden over 22 verschillende onderwerpen. Deze peilingen vonden plaats in resp. 1987, 1992, 1997, 2004 en 2011 (Wijnstra, 1988; Bokhove, Van der Schoot & Eggen, 1996; Janssen, Van der Schoot, Hemker & Verhelst, 1999; Janssen Van der Schoot & Hemker, 2005; Scheltens, Hemker & Vermeulen, 2013).

Hoewel er in de loop der tijd wel enige wijzigingen zijn opgetreden in aantal en inhoud van deze DBE-en, kan er voor 22 eenheden worden vergeleken over de tijd. Ten behoeve van de duidelijkheid in de weergave hebben we deze verdeeld in vier domeinen en drie figuren. De verdeling is weergegeven in afbeelding 1.

GETALLEN & BEWERKINGEN	VERHOUDINGEN, BREUKEN EN PROCENTEN	METEN EN VERBANDEN ¹
1. Getallen & getalsrelaties 2. Basisoperaties +/- 3. Basisoperaties x/: 4. Hoofdrekenen +/- 5. Hoofdrekenen x/: 6. Schattend rekenen 7. Bewerkingen +/- 8. Bewerkingen x/: 9. Samengestelde bewerkingen 10. Zakrekenmachine	11. Verhoudingen 12. Breuken 13. Procenten	14. Lengte 15. Oppervlakte 16. Inhoud 17. Gewicht 18. Toepassingen 20. Tijd 21. Geld 22. Verbanden

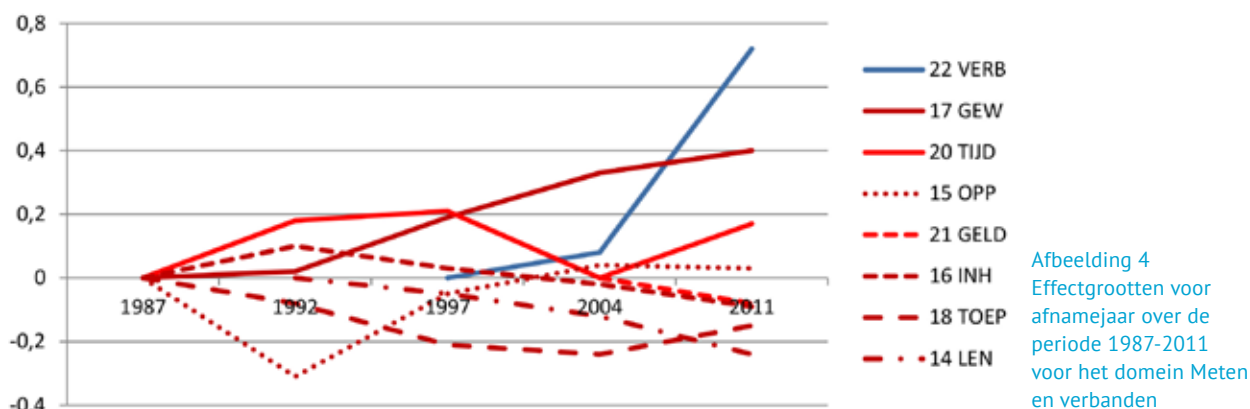
Afbeelding 1
Verdeling van
didactisch-
betekenisvolle
eenheden over
domeinen

In de afbeeldingen 2 en 3 brengen we het verloop in de tijd in beeld met behulp van de zogenaamde effect-schattingen. Bij PPON worden de uitkomsten doorgaans weergegeven op een vaste schaal met als gemiddelde 250 en als standaardafwijking 50. Die schaal is over de tijd constant, zodat we de veranderingen in schaalscores kunnen weergeven. Of een verandering betekenisvol is kan echter beter met behulp van effect-



schattingen worden weergegeven. Daarbij wordt het verschil in gemiddelde schaalscore gedeeld door de gemeenschappelijke standaarddeviatie van de beide verdelingen uitgedrukt in een positieve of negatieve waarde. Daarbij geldt het volgende waarderingskader (zowel positief als negatief): 0 tot 0,2 - verwaarloosbaar, geen betekenis; 0,2 tot 0,5 - klein effect; 0,5 tot 0,8 - matig effect; groter dan 0,8 - groot effect (Cohen, 1988).

In afbeelding 2 zien we grote en matige effecten, zowel positief als negatief. Positief blijken vooral Schattend rekenen, Getallen en getalsrelaties, Hoofdrekenen: optellen en aftrekken en ook wel Rekenen met de zakrekenmachine. Duidelijk negatief is de ontwikkeling bij Bewerkingen: vermenigvuldigen en delen, Samengestelde bewerkingen, Bewerkingen: optellen en aftrekken. Voor de overige DBE-en zijn de effectgroottes klein, dan wel betekenisloos. Dat geldt dus voor de twee schalen van Basisoperaties en voor Hoofdrekenen. In afbeelding 3 zien we een matig positief effect voor Procenten, met een constant stijgende lijn vanaf 1987. Bij de andere twee DBE-en, Breuken en Verhoudingen zijn geen betekenisvolle effectgroottes te zien. Die stijgen vanaf 1997 dus niet meer mee. In afbeelding 4 zijn twee domeinen gecombineerd. Wat onmiddellijk opvalt, is dat de schaal Verbanden vergeleken met 2004 een groot positief effect laat zien, terwijl dat bij de andere DBE-en in het domein Meten, meetkunde, tijd en geld, meestal betekenisloos blijft. Alleen de eenheid Gewicht vertoont, zeker vergeleken met 1987, een duidelijk, maar klein effect.

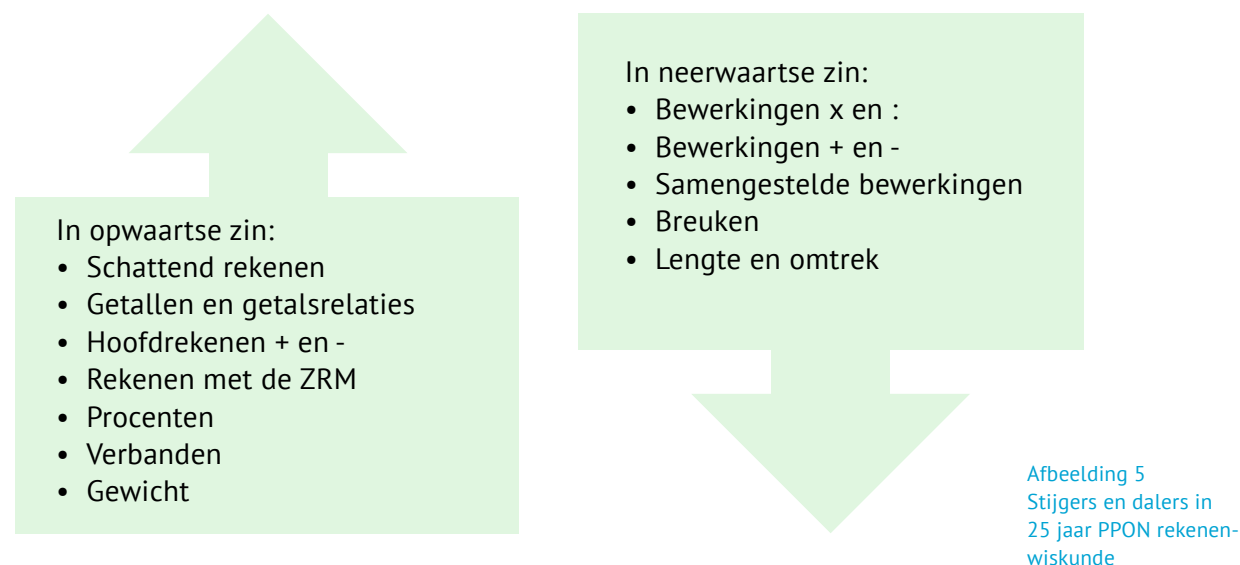


Afbeelding 4
Effectgrootten voor
afnamejaar over de
periode 1987-2011
voor het domein Meten
en verbanden

Deze grafieken leden tot een aantal algemene observaties. Ten eerste kan gesteld worden dat er duidelijke verschillen in ontwikkeling zijn als we de DBE-en onderling vergelijken. Er zijn eenheden die zich positief ontwikkelen, eenheden die ongeveer op hetzelfde niveau blijven en eenheden die duidelijk een negatieve ontwikkeling doormaken. Daarmee wordt helder gemaakt dat de prestaties in rekenen-wiskunde in groep 8 niet over een kam geschoren kunnen worden.

Ten tweede wordt met deze grafische weergave duidelijk dat er een flinke dynamiek in de leerprestaties is gezien over 25 jaar. Er zijn eenheden waarbij sprake is van ups en downs over de tijd: stijging wordt gevolgd door daling of andersom. Voorbeelden hiervan zijn de eenheden Getallen en getalsrelaties, Verhoudingen, Breuken, Tijd, Oppervlakte en Toepassingen.

Ten derde zien we zowel evidente stijgers als evidente dalers. Dat zijn eenheden waar vanaf 1987 een gestage ontwikkeling opwaarts, dan wel neerwaarts kan worden gesignaleerd. In Afbeelding 5 hebben we deze stijgers en dalers naast elkaar gezet.



Afbeelding 5
Stijgers en dalers in
25 jaar PPON rekenen-
wiskunde

Men kan de vraag stellen of de neergang op de ene eenheid, gecompenseerd wordt door de vooruitgang op andere eenheden. Ook is er de vraag of deze twee bewegingen elkaar nu in evenwicht houden wat het algemene rekenniveau betreft of dat er sprake is van een stijging, dan wel daling van het algemene reken-niveau. In het advies 'Rekenonderwijs op de basisschool' wordt gesteld dat de prestaties vrij stabiel zijn gebleven, maar wordt tegelijkertijd geponeerd dat men positieve en negatieve ontwikkelingen niet tegen elkaar kan wegstrepen. De achteruitgang in vooral Bewerkingen mag volgens de auteurs niet worden gerelativeerd door te wijzen op de toename in het correct gebruik van de rekenmachine (KNAW, 2009, p. 11). Echter, hoe we deze ontwikkelingen moeten waarderen is niet echt het onderwerp van deze bijdrage. Waar we in deze bijdrage echter vooral de nadruk op willen leggen, is het gegeven dat deze discussie alleen mogelijk is geworden door het ontwerp van peilingsonderzoek met een redelijk aantal didactisch betekenisvolle eenheden. Zonder deze detaillering was de dynamiek in leerprestaties en het onderscheiden van stijgers en dalers niet mogelijk geweest.

Een andere kwestie is de keuze van eenheden in relatie tot de evaluatie van het algemene niveau van leerprestaties, want bij het ontwerp van een peilingsonderzoek wordt het onderwijsleergebied verkaveld

in een reeks van DBE-en en wordt een gewicht gegeven aan de verschillende eenheden. Op dat moment wordt al een keuze gemaakt die bepalend is voor de evaluatie van het niveau, immers afhankelijk van de accentuering kan de conclusie luiden dat er sprake is van een stijging, dan wel een daling. Stel dat het accent in de opgavenverzameling van PPON was gelegd op de eenheden Getallen en getalsrelaties, Schattend rekenen en Hoofdrekenen, dan zou de conclusie zijn geweest dat het niveau was gestegen. Was echter de nadruk gelegd op de diverse vormen van Bewerkingen (met papier), dan zou de conclusie zijn geweest, dat het niveau was gedaald².

ANDERE SYSTEEMEVALUATIES VAN HET NIVEAU REKENEN-WISKUNDE

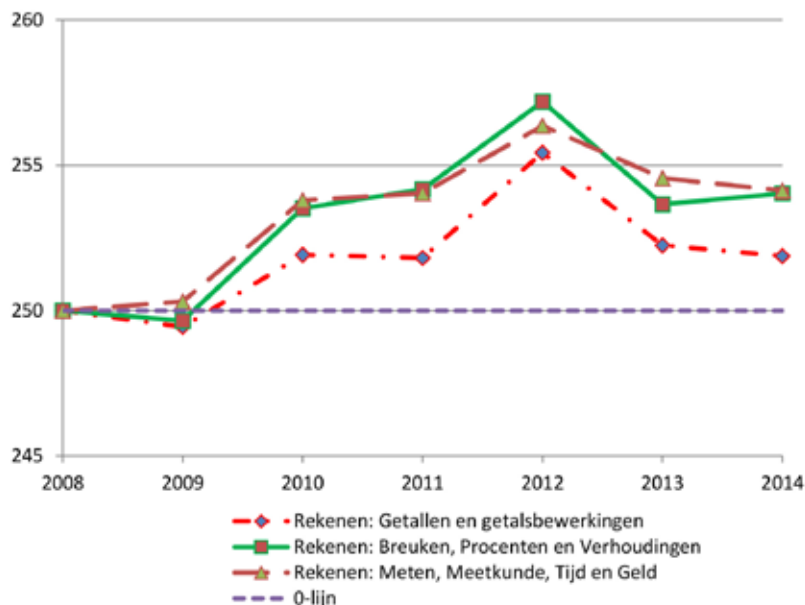
Er zijn ook andere onderzoeken die informatie kunnen leveren over de vraag of er sprake is van een algemene stijging, daling, dan wel stabilisering van niveau bij rekenen-wiskunde in Nederland. We kunnen te rade gaan bij andere metingen van het onderwijsniveau op systeemniveau. Er zijn drie systeemevaluaties die zich daartoe goed lenen, te weten de jaarlijkse peiling van het onderwijsniveau (JPON), een variant van PPON, en twee internationale peilingsonderzoeken, TIMSS (Trends in International Measurement and Science Study) en PISA (Program of International Student Assessment). Uitkomsten gebaseerd op toetsen van het leerlingvolgsysteem laten we hierbij buiten beschouwing, omdat deze in principe niet ontwikkeld zijn voor het doel van nationale evaluatie. Om die reden vervalt ook de vergelijking met de uitkomsten van het project COOL 3-15, immers in dat project wordt de vaardigheid in rekenen-wiskunde ook gemeten met toetsen afkomstig uit het leerlingvolgsysteem (Driessen e.a., 2009).

JPON: Jaarlijkse Peiling van het Onderwijsniveau³

Sinds 2008 wordt door Cito als variant op PPON een peiling uitgevoerd die gebruik maakt van de landelijk representatieve meting in jaargroep 8 van de Eindtoets basisonderwijs (EB). In deze toets kunnen vanwege de beschikbare toetstijd minder opgaven voor rekenen-wiskunde worden opgenomen dan in een peiling van PPON. Dat betekent dat er ook minder rapportage-eenheden gedefinieerd kunnen worden waarop betrouwbaar en valide kan worden gerapporteerd. In de afgelopen periode waren dat er drie, en met de invoering van de referentieniveaus voor rekenen worden dat er vanaf 2014 vier. De EB heeft een ander doel dan systeemevaluatie en leent zich eigenlijk minder goed voor een analyse overeenkomstig die van PPON⁴. Toch bleek het ook hier mogelijk om een vaste schaal te construeren, met een gemiddelde van 250 en een SD van 50, met als basisjaar 2008 (zie verder Hemker & Van Weerden, 2009). Doordat deze meting elk jaar opnieuw wordt uitgevoerd kan een trend in beeld gebracht worden over zeven opeenvolgende jaren voor drie didactisch betekenisvolle eenheden:

1. Getallen en getalsbewerkingen
2. Breuken, procenten en verhoudingen
3. Meten, meetkunde, tijd en geld

De trend, in beeld gebracht in afbeelding 6, laat zien dat er vanaf 2009 sprake is van een stijging voor alle drie de eenheden, dat er na een stabilisatie van 2010 naar 2011, in 2012 opnieuw vooruitgang wordt geboekt, maar dat daarna het niveau weer wat zakt, om te stabiliseren in 2014. Het verschil tussen 2008 en 2014 blijft wel duidelijk positief, waarbij er een klein onderscheid is tussen Getallen en getalsbewerkingen enerzijds en de andere twee eenheden anderzijds (zie verder Hemker & Van Weerden, 2015). Opvallend is verder dat de trend bij de drie eenheden min of meer hetzelfde verloop hebben. Het onderscheid tussen de drie rapportage-eenheden in het verloop van de trend is verwaarloosbaar. Omdat de periodiciteit bij PPON een andere is, de laatste meting was in 2011 en die daarvoor in 2004, is een rechtstreeks vergelijking niet te maken. Wel is duidelijk dat van de dynamiek in DBE-en hier geen sprake is. De drie gehanteerde eenheden zijn minder gedetailleerd uitgewerkt in de EB en omvatten daardoor meerdere DBE-en die bij PPON worden onderscheiden. Dat maakt een inhoudelijke interpretatie van de trends, zoals in PPON gebeurt, moeilijk, zo niet onmogelijk.⁵



Afbeelding 6
Trends in de resultaten
voor 2008 tot en met 2014
voor Rekenen bij JPON

TIMSS - Trends in International Mathematics en Science Study

Over het internationale onderzoek bij leerlingen in jaargroep 6 van TIMSS werd in 2011 voor de vierde keer verslag gedaan. Daaruit bleek dat de Nederlandse leerlingen er in absolute zin op vooruit zijn gegaan sinds 2007, maar dat zij in rangorde bij rekenen-wiskunde van de vijfde plaats naar de achtste zijn gezakt, terwijl ze bij de eerste meting in 1995 nog op de vierde plaats hebben gestaan en in 2013 op de zesde. Vergelijken met 2003 wordt in 2011 dezelfde score behaald: 540 op een schaal met als gemiddelde 500 (zie verder Meelissen e.a., 2012). Bij TIMSS maakt men onderscheid in drie inhoudelijke domeinen en drie cognitieve domeinen, die met elkaar gekruist, een matrix opleveren waarin de 175 opgaven zijn geplaatst (zie afbeelding 7). Opvallend is nu dat bij de jaarvergelijking over de domeinen bij Gegevensweergave een duidelijke plus wordt gerapporteerd (14 punten op de schaal, significant). Bij de andere eenheden zien we geen significante verschillen. Ook hier blijkt dat het onderscheid naar eenheden meer zicht geeft op de uitkomsten. Tegelijkertijd blijkt hieruit geen verschil in uitkomsten tussen bijvoorbeeld bewerkingen en schattend rekenen, omdat dat verschil in DBE-en hier niet gemaakt wordt.

Inhoudelijke domein	Cognitieve domein			Opgaven in %	Leerstofgebieden
	weten	toepassen	redeneren		
Getallen	38	33	17	50	8
Geometrische vormen en meten	24	26	11	35	7
Gegevensweergave	9	12	6	15	3
Opgaven in %	40	41	19	100	

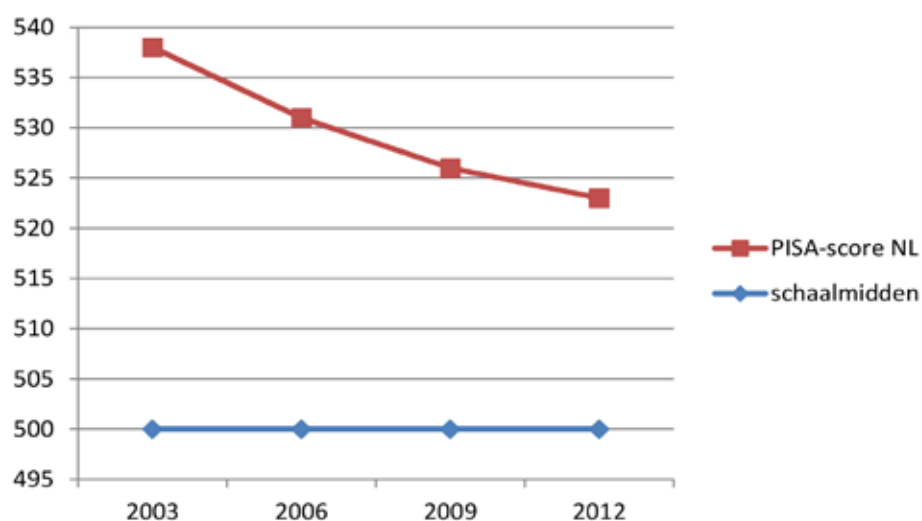
Afbeelding 7 Aantal opgaven per eenheid Rekenen in TIMSS 2011

	2007	2011	2011-2007
Inhoudelijke domein			
Getallen	539	543	4
Geometrische vormen en meten	522	524	2
Gegevensweergave	545	559	14
Cognitieve domein			
weten	528	537	9
toepassen	540	540	0
redeneren	537	543	6

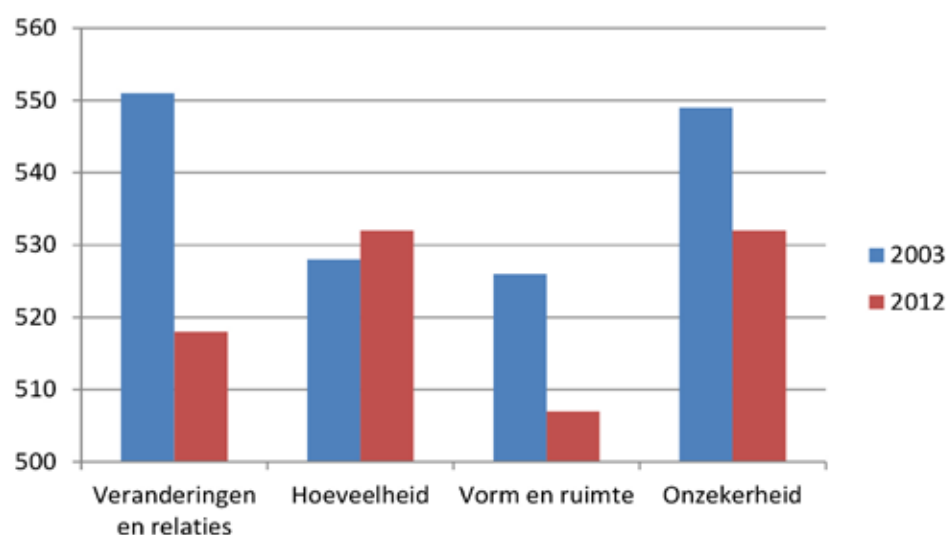
Afbeelding 8
Uitkomsten op TIMSS
2007-2011 naar domein/
meeteenheid

PISA - Program for International Student Assessment

In 2012 was wiskunde - net als in 2003 - het hoofdonderwerp van dit internationaal vergelijkend onderzoek, bij 15-jarigen, waar 64 landen aan deelnamen. Nederland doet mee vanaf 2003 en om de drie jaar wordt er een peiling verricht (zie verder Kordes e.a., 2013). De resultaten zijn weergegeven in afbeelding 9. De trend laat een duidelijk neergaande lijn zien. Ook PISA gebruikt een vaste schaal, net al PPON, maar met als gemiddelde 500. In afbeelding 10 zijn de vier onderscheiden rapportage-eenheden weergegeven: Veranderingen & relaties, Hoeveelheid, Vorm en ruimte en Onzekerheid. Kort door de bocht geformuleerd staan deze labels voor respectievelijk: algebra, rekenkunde, meetkunde en statistiek/kansrekening. Als we de uitsplitsing in deze eenheden in beeld brengen bij trend van 2003 naar 2012, dan blijkt dat er voor drie van de vier eenheden sprake is van een verslechtering, maar juist niet bij de eenheid Hoeveelheden. En die staat van de vier nog het meest voor wat we hier onder rekenen-wiskunde verstaan. De andere eenheden gaan veel meer over onderwerpen die meer in het VO dan op de basisschool aan de orde zijn. We kunnen op basis van deze resultaten dus niet concluderen dat leerprestaties bij rekenen achteruit zijn gegaan.



Afbeelding 9
PISA-scores Nederland van
2003 tot en met 2012



Afbeelding 10
Vergelijking van de
PISA-scores Nederland
voor 2003 en 2012 op de
subschalen

Conclusie en discussie

Met deze uitkomsten op het PISA-onderzoek is goed te illustreren hoe belangrijk het onderscheid naar rapportage-eenheden is bij peilingsonderzoek. Kijken we alleen naar de algemene trend voor wiskunde zoals die wordt geschetst in afbeelding 9, dan zou de conclusie kunnen zijn dat er sprake is van een geleidelijke achteruitgang. Kijken we echter meer gedetailleerd, zoals in afbeelding 10, dan blijkt de daling bij de rapportage-eenheid Hoeveelheden zich niet voor te doen, een schaal die nu juist vooral betrekking heeft op wat we als basaal rekenen kunnen aanduiden.

Let wel: we kunnen de uitkomsten van deze internationale peilingsonderzoeken niet zomaar vergelijken

met PPON en JPON⁶. Het gaat ook om leerlingen in andere jaargroepen, dus met andere leeftijden. Wel laten deze onderzoeken goed zien dat het belangrijk is verder te kijken dan alleen een algemene indicatie betreffende het niveau van het rekenonderwijs. Juist de detaillering maakt het mogelijk meer gedegen conclusies te trekken. Vergeleken met het grote aantal didactisch betekenisvolle eenheden dat wordt gehanteerd bij PPON blijken de categorisering van zowel JPON, als van TIMSS en PISA maar weinig relevante inhoudelijke conclusies toe te laten. Het accent ligt bij de internationaal vergelijkende studies sterk op de onderlinge positie van de deelnemende landen en veel minder op gedifferentieerde inhoudelijke feedback op het behaalde niveau. Het onderscheid naar drie of vier domeinen/schalen leent zich daar ook niet goed voor. Bij JPON zien we nauwelijks verschil tussen de drie rapportage-eenheden. Bij TIMSS zien we bij de zes eenheden één significant verschil bij de vergelijking over de jaren. Bij PISA worden domeinen gedefinieerd met een grote reikwijdte en 'Rekenkunde' onderscheidt zich daarbij juist positief. Vergelijken we deze uitkomsten echter met de 22 DBE-en bij PPON dan zien we niets terug van de grote dynamiek in inhoudelijke ontwikkelingen die we daar waarnemen. De achteruitgang van beweringen wordt niet gesignaleerd, evenmin als de sterke vooruitgang van schattend rekenen en rekenen met verbanden.

CONCLUSIES

Als we de resultaten van de rekenpeilingen van PPON over de afgelopen 25 jaar overzien kunnen we de volgende conclusies trekken:

- Er is in de peilingen van PPON een forse dynamiek waargenomen in de ontwikkeling van de rekenvaardigheid.
- Om die dynamiek te kunnen waarnemen is het cruciaal welke rapportage-eenheden worden gekozen. Of het niveau stijgt dan wel daalt, kan niet eenduidig geconcludeerd worden.
- Samenvoegen van rapportage-eenheden verdoezelt de eventueel aanwezige effecten.

Een 'evenwichtig' samengestelde rekentoets zou over de afgelopen 25 jaar ten onrechte tot de conclusie hebben kunnen leiden dat er in het rekenvaardigheidsniveau van leerlingen nauwelijks of geen verandering is opgetreden. In deze bijdrage en ook die van anderen, (o.a. Van Putten & Hickendorff, 2006) wordt duidelijk gemaakt dat er nu een discussie mogelijk is over het niveau, juist omdat er bij PPON in de huidige opzet een fijnmazige meting wordt verricht.

Peilen ten behoeve van de bepaling van het nationaal rekenniveau is vooral zinvol als dat gebeurt aan de hand van een fijnmazige domeinbeschrijving, waarbij voldoende rapportageschalen worden gedefinieerd. Dat levert een gedetailleerd en goed uitgebalanceerd beeld op over de vaardigheden van de leerlingen. Cruciaal daarbij is de samenstelling van de opgavenverzameling, de keuze van DBE-en (de rapportageschalen) en de onderlinge verhoudingen tussen die rapportage-eenheden. In wezen is dit een vraag betreffende de validiteit van het meetinstrument. Net als bij de bepaling van een goede toets, waarin de bedoelde inhoud correct vertegenwoordigd moet zijn in de opgaven, geldt dat ook op grootschaliger niveau, namelijk voor het instrumentarium van een geheel peilingsonderzoek⁷.

PPON is een onderzoek dat vanuit de overheid wordt gefinancierd. Opdrachtgever is OCW, die in de afgelopen 25 jaar Cito met de planning en uitvoering heeft belast. Door een verandering in de Wet SLOA wordt met ingang van 2014 de regie voor systeemevaluatie zoals gerealiseerd door PPON, bij de Inspectie van het onderwijs neergelegd.

Opdrachtgevers zijn dikwijls geneigd te kiezen voor een gemakkelijker en vaak ook goedkoper geachte oplossing. Dat is in dit kader bijvoorbeeld de variant waarin JPON als vervanger zou worden gezien van PPON of een variant waarin wordt aangestuurd op het vervangen van een afzonderlijke nationale peiling, door het verzamelen van gegevens uit reeds in gebruik zijnde leerlingvolgsystemen.

Beide alternatieven hebben zo hun bezwaren. Het eerste alternatief is met bovenstaand exposé genoegzaam onderbouwd. Het tweede alternatief is hier nog niet besproken, maar levert ook de nodige vragen op. Zonder op deze plek uit te wijden kan gesteld worden dat leerlingvolgsystemen voor Rekenen-wiskunde, zoals we die nu kennen, geen gedetailleerde opgavenverzameling kennen, vergelijkbaar met PPON en hoogstens een rapportage kunnen opleveren op de vier domeinen van het referentiekader. Die situatie is beschreven in het JPON onderzoek, waarin ook jaargroep 4 werd geanalyseerd (Van Weerden, Hemker & Mulder, 2014). Daar komt nog bij dat men in dat geval afhankelijk is van gegevens die door de scholen zelf worden verzameld en geregistreerd. Bovendien geldt ook hier dat de opgaven niet geselecteerd zijn voor het gebruik in een peilingsonderzoek en dat is een wezenlijk verschil.

Wat is het toekomstperspectief voor de peilingsonderzoeken rekenen-wiskunde?

Volgens de planning van PPON zou er in 2017 weer een peilingsonderzoek moeten plaatsvinden in jaargroep 8. Of en hoe dat zal gaan plaatsvinden is nu (eind 2015) nog niet bekend. Helemaal onzeker is hoe het verder gaat met het peilingsonderzoek voor groep 5 en de eindgroep van het sbo. Een peiling voor groep 5 zou al in 2016 aan de orde zijn en die in het sbo in 2019 of 2020.

Er zijn geen signalen dat voor die groepen de continuïteit wordt gehandhaafd. De peiling voor groep 8 zou in ieder geval gecontinueerd worden in de JPON-variant. Dat zou betekenen dat er geen verzameling meer zou zijn van gekoppelde gegevens over het onderwijsaanbod van de betrokken scholen. Een

variant die ook perspectief kan bieden is een koppeling aan de adaptieve centrale eindtoets die in moet gaan vanaf 2018. Er is dan sprake van een digitale toets, die niet voor alle leerlingen hetzelfde zou hoeven te zijn en waarmee dan ook een grotere opgavenverzameling dan nu mogelijk is kan worden ingezet. Dat geeft rijkere rapportagemogelijkheden.

TOT SLOT

De bovenstaande tekst is als lezing uitgesproken tijdens de Panamaconferentie van januari 2015. Deze bijdrage is toen gecombineerd met een workshop, waarin de deelnemers op een aantal vragen konden reageren. Zo was er een vraag wat men zou willen veranderen aan de opzet, uitvoering rapportage van PPON. Opvallend was dat de deelnemers overwegend reageerden in de sfeer van 'niets veranderen' en 'vooral zo door gaan'. Opvallend was ook dat bij deze groep niemand aandrong op een rapportage direct aan de deelnemende scholen, iets dat in de gebruikelijke opzet niet kan worden geleverd, maar vaak door externe critici wordt geopperd als belangrijke mogelijke verbetering. De belangrijkste verbeteringsuggestie was gericht op de concrete bruikbaarheid van de resultaten voor leerkrachten. Graag zou men zien dat de uitkomsten vaker vertaald zouden kunnen worden in suggesties voor aanpassing en verbetering van de onderwijspraktijk, iets wat tot nu sporadisch is gerealiseerd in het kader van aanvullende onderzoeken, bijvoorbeeld naar rekenstrategieën en oplossingsmethoden, zoals die samen met de Universiteit Leiden zijn uitgevoerd (Fagginger Auer e.a., 2015). Wat deze deelnemers betreft zou er vooral doorgegaan moeten worden op de ingeslagen weg, met een intensivering van deze aanvullende en meer op de praktijk gerichte onderzoeken.

In hoeverre dat bij peilingsonderzoek in Nederland nog mogelijk is zal in de toekomst moeten blijken.

Literatuur

- Bock, R.D., Mislevy, R.J. & Woodson, C. (1982) The next stage in educational assessment. *Educational Researcher*, 11, (3), 4, 16.
- Bokhove, J., van der Schoot, F. & Eggen, T. (1996). *Balans van het rekenonderwijs aan het einde van de basisschool 2. Uitkomsten van de tweede peiling rekenen/wiskunde einde basisonderwijs*. PPON-reeks nr. 8a. Arnhem: Cito Instituut voor Toetsontwikkeling.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (revised edition). Hillsdale, NJ: Lawrence Erlbaum.
- Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J., & Van der Veen, I. (2009). *Cohortonderzoek COOL 5-18. Technisch rapport basisonderwijs, eerste meting 2007/08*. Nijmegen/Amsterdam, ITS/Kohnstamm Instituut.
- Expertgroep Doorlopende Leerlijnen (2008). *Over de drempels met taal en rekenen*. Eindrapport. Enschede, SLO.
- Fagginger Auer M.F., Hickendorff, M. & Van Putten, C.M. (2015). *Strategiegebruik bij rekenen afleiden uit het schriftelijk werk van basisschoolleerlingen*. *Pedagogische Studiën*, 92(1), 9-23.
- Hemker, B.T. & Van Weerden, J.J., (2009). *Peiling van de rekenvaardigheid en de taalvaardigheid in groep 8 en jaargroep 4 in 2008*. Cito, Arnhem.
- Hemker, B.T. & Van Weerden, J.J., (2015). *Peiling van de rekenvaardigheid en de taalvaardigheid in groep 8 van het basisonderwijs in 2014*. Cito: Arnhem.
- Janssen, J., Van der Schoot, F., Hemker, B., & Verhelst, N. (1999). *Balans van het reken-wiskunde-onderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 1997*. PPON-reeks nr. 13. Arnhem: Cito Instituut voor Toetsontwikkeling.
- Janssen J., Van der Schoot, F. Hemker, B. (2005). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 4. Uitkomsten van de vierde peiling in 2004*. PPON-reeks nr. 32. Arnhem: Citogroep.
- Koninklijke Nederlandse Akademie van Wetenschappen (KNAW). (2009) *Advies KNAW-commissie rekenonderwijs basisschool. Analyse en sleutels tot verbetering*. Amsterdam, KNAW.
- Kordes, J., Bolsinova, M., Limpens, G. & Stolwijk, R. (2013). *Resultaten PISA-2012. Praktische kennis en vaardigheden van 15-jarigen*. Arnhem, Cito.
- Meelissen, M.R.M., Netten, A., Drent, M., Punter, R.A., Droop, M., Verhoeven, L. (2012) *PIRLS- en TIMSS-2011. Trends in leerprestaties in Lezen, Rekenen en Natuuronderwijs*. Nijmegen: Radboud Universiteit: Enschede: Universiteit Twente.
- Scheltens, F. B. Hemker, J. Vermeulen (2013). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 5. Uitkomsten van de vijfde peiling in 2011*. PPON-reeks nr. 51. Arnhem, Cito.
- Schoot, F. van der, (2008). *Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPON*. Arnhem: Cito Instituut voor Toetsontwikkeling.
- Van Putten, C.M. & Hickendorff, M. (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.
- Van Weerden, J. & Hiddink, L. (2013). *Balans van het basisonderwijs. PPON: 25 jaar kwaliteit in beeld*. Arnhem: Cito Instituut voor Toetsontwikkeling.
- Van Weerden, J.J., Hemker, B.T.H., & Mulder, K. (2014). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2013*. Cito: Arnhem.
- Wijnstra, J.M. (red.) (1988). *Balans van het rekenonderwijs in de basisschool. Uitkomsten van de eerste rekenpeiling medio en einde basisonderwijs*. PPON-reeks nr. 1. Arnhem: Instituut voor Toetsontwikkeling.

The National Assessment (PPON) exists over 25 years. The aims of this assessment is to offer an empiric fundament for discussion over content, quality and level of education.

In this study we emphasize the importance of the design of PPON. Hence we pay attention to the lessons learned, based on the results, for the design of National Assessment in the future. In this study there is a comparison between the design of PPON and other large scale assessments as TIMSS and PISA. In the end we take a look at the future of PPON.

The conclusions below can be drawn based on the results of the mathematics assessments of PPON in the last 25 years:

- *The dynamism in the mathematics ability was huge.*
- *To observe this dynamism it's crucial to choose the correct detailed design.*
- *Effects will be covered up when a less detailed design is administered.*

The ground discussion on the mathematics ability in the Netherlands was only made possible by the detailed design of PPON.

Noten

¹ Verbanden is hier, afwijkend van de indeling bij de referentieniveaus, vanwege de mogelijkheid tot vergelijking over de tijd niet als afzonderlijk domein opgevoerd en omvat slechts één DBE. Meetkunde als DBE ontbreekt, vanwege een slechte fit in de analyses

² Er zijn diverse mogelijkheden om rekenopgaven te verkavelen over eenheden en ook diverse mogelijkheden om de verschillende relatieve bijdragen van de eenheden te verdisconteren in een algemene maat. Die discussie laten we hier even rusten.

³ In de loop der tijd is dit project ook de Jaarlijkse Meting Taal en Rekenen genoemd, om eventuele verwarring met PPON te voorkomen en te benadrukken dat het alleen om deze twee onderwijsleergebieden gaat.

⁴ De andere functie van de toets heeft o.a. effect op de keuze van opgaven, zoals de spreiding van de moeilijkheidsgraad en de gemiddelde moeilijkheidsgraad. Verder wordt bij peilingsonderzoek met deels nieuwe en deels geheim gehouden opgaven gemeten. Bij de EB worden alle opgaven na afname gepubliceerd.

⁵ Om over een afzonderlijke eenheid te kunnen rapporteren zijn er voldoende opgaven nodig die daarbinnen kunnen worden gerubriceerd. Bij een toets met 60 rekenopgaven zijn 3 tot 4 eenheden wel het maximum.

⁶ Ten eerste is er sprake van nationale versus internationale onderzoeken. Ten tweede gaat het om leerlingen van verschillende leeftijden. Verder zijn de uitgangspunten voor instrumentconstructie vaak anders.

⁷ Naast voldoende opgaven over de verschillende onderwerpen is het overigens ook belangrijk het type opgaven voldoende te variëren. Het gaat te ver om dat hier afzonderlijk uit te werken.