



Drie onderzoeken naar reken-wiskundeonderwijs

J. Bokhove

In de jaren 2002 en 2003 vond een drietal onderzoeken plaats naar het niveau van het reken-wiskundeonderwijs: PPON, TIMSS en PISA. In dit artikel beschrijf ik eerst de opzet van de drie onderzoeken. Daarna wordt ingegaan op de gebruikte instrumenten en ten slotte worden de resultaten van de onderzoeken in relatie met elkaar beschreven. Ik sluit dit artikel af met enkele vragen, opmerkingen en kanttekeningen.

1 Aard en opzet van de onderzoeken

De drie onderzoeken PPON, TIMSS en PISA beschrijven opbrengsten van reken-wiskundeonderwijs. Het zijn beschrijvende en geen verklarende onderzoeken. Bij TIMSS en PISA gaat het om vergelijkingen van deze opbrengsten met die van andere landen en het vergelijken met uitkomsten van vorige onderzoeken. In het PPON-onderzoek gaat het om trends in ons Nederlandse reken-wiskundeonderwijs, door vergelijkingen te maken met vorige onderzoeken. Maar het gaat in PPON ook om een inhoudelijke beschrijving van wat leerlingen in het reken-wiskundeonderwijs presteren. Experts geven bovendien hun oordeel of het behaalde en gewenste niveau overeenkomen.

In de volgende paragrafen beschrijf ik eerst de drie afzonderlijke onderzoeken qua aard en opzet en sluit af met het vergelijken van de onderzoeken, kijkend naar overeenkomsten en verschillen.

PPON

PPON staat voor 'Periodieke Peiling van het OnderwijsNiveau'. Het is een door het Cito, in opdracht van de minister van Onderwijs, uitgevoerd project. Het onderzoek betreft verschillende vakgebieden. Voor rekenen-wiskunde wordt het onderwijs op twee momenten in het basisonderwijs onderzocht, medio (in groep 5) en aan het eind van de basisschool (in groep 8). Verder wordt onderzoek gedaan in het speciaal onderwijs. Voor rekenen-wiskunde vond de eerste mediopeiling plaats in 1987. Daarna volgden onderzoeken in 1992 en 1997. In januari-februari 2003 is in jaargroep 5 het vierde peilingsonderzoek halverwege de basisschool uitgevoerd. De resultaten daarvan zijn in januari 2005 gepubliceerd en toege-

licht op de Panama-conferentie in Noordwijkerhout.

Het gaat bij PPON om een steekproefonderzoek. Bij het trekken van de steekproef is uitgegaan van drie groepen scholen, op basis van zogenoemde leerlinggewichten:

- scholen met een leerlingpopulatie die overwegend uit kinderen van ouders met een afgeronde voortgezette opleiding bestaat en met weinig allochtone leerlingen;
- scholen met relatief meer Nederlandse arbeiderskinderen en weinig allochtone leerlingen;
- scholen met vooral Nederlandse arbeiderskinderen en allochtone leerlingen.

In verhouding tot de omvang van iedere groep scholen is een basissteekproef getrokken en voor elke school uit de basissteekproef nog een aantal reservescholen. Dat laatste is nodig omdat deelname aan het onderzoek op vrijwillige basis geschiedt. Omdat het onderzoek in 2003 niet in het begin van groep 5 plaats vond - in tegenstelling tot vorige peilingen - maar in de maanden januari-februari, dus halverwege groep 5, was aanvullend onderzoek nodig om de uitkomsten van de peiling 2003 te kunnen vergelijken met die van vorige peilingen.

Hoe gaat de afname van de toets? Leerlingen uit eenzelfde klas krijgen verschillende boekjes met opgaven voorgelegd. Een dergelijke werkwijze zorgt ervoor dat met een gering aantal scholen in de peiling toch een goed beeld van de situatie kan worden verkregen. Naast de door de leerlingen te maken toetsen is er een vragenlijst, die ingevuld wordt door de leraren van de groepen 3, 4 en 5, waarin vooral gevraagd werd naar het onderwijsaanbod. Tevens worden achtergrondkenmerken van de leerlingen geïnventariseerd door middel van een leerlingenlijst. Voor de afname worden proefleid(st)ers ingeschakeld, om te waarborgen dat de toetsen overal op dezelfde manier worden afgenomen.

TIMSS

TIMSS staat voor 'Trends in International Mathematics and Science Study'. Het is een internationaal onderzoek onder auspiciën van de IEA, de 'International Association for the Evaluation of Educational Achievement'. De IEA is een onafhankelijke organisatie, waarin onderzoeksinstituten uit meer dan vijftig landen samenwerken bij het verrichten van internationaal vergelijkende studies die tot doel hebben opbrengsten en contexten van onderwijs in kaart te brengen. In 1995 is voor het eerst in Nederland een TIMSS-onderzoek uitgevoerd. Dit gebeurde in drie populaties, namelijk in groep 5 en 6 van het basisonderwijs, in de eerste twee leerjaren van het voortgezet onderwijs en in het laatste jaar van het voortgezet onderwijs. In 1999 is het onderzoek in het tweede leerjaar van het voortgezet onderwijs herhaald. In 2003 zijn data verzameld in groep 6 van het basisonderwijs en leerjaar 2 van het voortgezet onderwijs. In december 2004 verscheen het rapport 'TIMSS-2003 Nederland', over leerprestaties in exacte vakken in het basisonderwijs.

Bij de steekproeftrekking is men bij de TIMSS-onderzoeken op dezelfde manier te werk gegaan als bij PPON. Het gaat om een gestratificeerde steekproef van scholen naar gemiddeld leerlinggewicht. Voor iedere school in de basissteekproef werden twee reservescholen getrokken. De IEA stelt als eis voor een goede steekproef dat er bij de basissteekproef een respons van minimaal 50 procent is (dat wil zeggen dat minstens 50 procent van de aangeschreven scholen zich bereid verklaart aan het onderzoek mee te doen) en dat in totaal, met de reservescholen meegerekend, er een respons van minimaal 85 procent is. De steekproef in Nederland voldeed met percentages van respectievelijk 52 en 87 aan die eis. De steekproef bij TIMSS was, net als bij PPON, zodanig dat verschillende groepen scholen goed vertegenwoordigd waren.

De toetsen werden in de meeste gevallen afgenomen door de leerkracht van groep 6, in de periode van de tweede helft van april tot de tweede helft van juni. Naast toetsgegevens zijn er ook achtergrondkenmerken van de deelnemende leerlingen en scholen verzameld.

PISA

PISA is de afkorting voor 'Programme for International Student Assessment'. Het PISA-onderzoek is geëntamerd door de Organisatie voor Economische Samenwerking en Ontwikkeling (OESO). Deze organisatie is van mening dat het onderwijsniveau een belangrijke indicator is van economische ontwikkeling. De OESO was oorspronkelijk een organisatie van westerse, geïndustrialiseerde landen, maar is nu, met dertig lidstaten, breder samengesteld. Aan het onderzoek in 2003 namen alle dertig lidstaten deel, uitgebreid met elf niet-lidstaten, de zogenaamde partnerlanden.

Het PISA-project wordt op internationaal niveau uitgevoerd door een samenwerkingsverband waarin de diverse

deelnemende landen vertegenwoordigd zijn. Voor Nederland maakt het Cito deel uit van dit samenwerkingsverband. Het Cito is door de minister dan ook aangewezen om de gegevens in Nederland te verzamelen.

PISA is gericht op vijftienjarige leerlingen, tegen het eind van de leerplichtige leeftijd en is opgezet als een driejaarlijks programma.

De eerste ronde vond plaats in 2000. Het accent lag in dat jaar op de leesvaardigheid van deze leerlingen. Wiskunde werd toen slechts in beperkte mate getoetst. In de tweede cyclus van 2003 lag het accent op vaardigheden in het toepassen van op school geleerde wiskunde, naast leesvaardigheid, kennis en vaardigheden in natuurwetenschappen en 'probleem oplossen'. Ik beperk me in dit artikel tot het onderdeel wiskunde.

Overeenkomsten en verschillen

In de drie onderzoeken worden aan leerlingen, in een steekproef van scholen, opgaven voorgelegd en worden er over scholen bovendien, middels vragenlijsten, aanvullende gegevens verzameld. Men is daarbij afhankelijk van de vrijwillige medewerking van scholen. Dit noodzaakt tot een werkwijze waarbij voor iedere school die in de eerste trekking zit, twee of meer reservescholen getrokken worden. Wil de eerst getrokken school z'n medewerking niet verlenen, dan kan een van de reservescholen in het onderzoek worden betrokken.

In de steekproeftrekking is een procedure gevolgd die rekening houdt met de grootte van de klas. Die kan in het voortgezet onderwijs namelijk nogal variëren.

Bij alle drie de onderzoeken zijn vergelijkingen met eerdere uitkomsten mogelijk, voor PISA slechts in beperkte mate, omdat het onderzoek naar wiskunde in 2000 beperkt was. Bovendien, omdat in het onderzoek van 2003, in tegenstelling tot dat van 2000, leerlingen uit het speciaal voortgezet onderwijs meededen. Voor PISA en TIMSS ligt het accent op de vergelijking van prestaties met die van leerlingen uit andere deelnemende landen.

2 Gebruikte instrumenten

PPON- peilinginstrumenten

De peilinginstrumenten van PPON bestaan uit een vragenlijst voor leraren over hun 'onderwijsaanbod', de 'toetsen' en een 'leerlingenlijst'.

De aanbodvragenlijst bevat vragen over:

- de gebruikte methode, oefenmaterialen en computerprogramma's;
- de gehanteerde differentiatievormen, zorgverbreding en remediëring;
- de algoritmen die in het cijferend rekenen worden aangeleerd.



figuur 1: de toetsopgaven 5, 6, 7 ‘verhoudingen’ uit het PPON-rapport

De toetsen bevatten overwegend open vragen en zijn verdeeld over elf onderwerpen. In het onderdeel ‘hoofdrekenen’ komen de basisoperaties voor optellen, aftrekken, vermenigvuldigen en delen aan bod.

Instrumenten bij TIMSS

De opgaven in de TIMSS-toets kennen een inhoudelijke dimensie en een cognitieve dimensie.

Inhoudelijke domeinen	Cognitieve domeinen					Aantal opgaven	Aandeel in %
	Feitenkennis en procedures	Concepten gebruiken	Routine-problemen oplossen	Redeneren			
Getallen	15	16	26	6	63	40	
Patronen, vergelijkingen en verbanden	3	5	9	7	24	15	
Meten	9	3	12	9	33	20	
Meetkunde	11	8	4	1	24	15	
Gegevens	0	5	9	3	17	10	
<i>Aantal opgaven</i>	38	37	60	26	161		
<i>Aandeel in %</i>	20	20	40	20		100	

figuur 2: aantal opgaven per cognitief domein bij het TIMSS-onderzoek

Schriftelijk zijn er toetsen voor de onderwerpen:

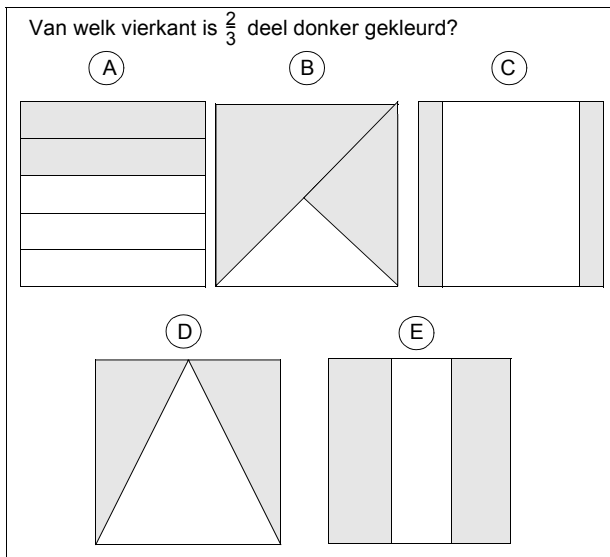
- getallen en getalrelaties;
- de bewerkingen optellen en aftrekken, vermenigvuldigen en delen en meer complexe bewerkingen;
- meten en meetkunde;
- tijd en geld;
- verhoudingen.

In figuur 1 ziet u enkele voorbeelden uit de toets verhoudingen.

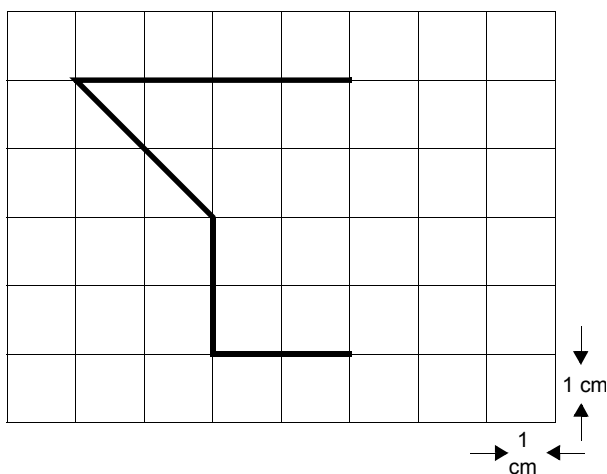
Met de leerlingenlijst, in te vullen door de meewerkende leraren van groep 3, 4 en 5, wordt informatie verzameld over sekse, formatiegewicht en leertijd (al of niet vertraagd).

In figuur 2 is weergegeven hoeveel opgaven er per inhoudelijke en cognitieve dimensie voor rekenen zijn opgenomen.¹ Deze toets bevat veel open opgaven, in ieder geval veel meer dan in de toets van 1995, hoewel het merendeel van de opgaven uit meerkeuzeopgaven bestaat. Het antwoord op een open vraag kan één, twee of drie punten opleveren. Een goed beantwoorde meerkeuzevraag levert één punt op.

Uit al die opgaven zijn twaalf toetsboekjes samengesteld, die steeds een verschillende combinatie van blokken met reken- en natuurkundeopgaven bevatten. Elk toetsboekje omvat minimaal twee rekenblokken. Elk rekenblok komt minimaal in twee toetsboekjes voor. In figuur 3 en 4 vindt u twee voorbeelden; de eerste is een meerkeuzeopgave, de tweede een open vraag uit het TIMSS-onderzoek.



figuur 3: meerkeuzeopgave in het domein getallen



De hokjes in het rooster hierboven hebben een oppervlakte van 1 vierkante centimeter. Maak de figuur zo af dat het een oppervlakte heeft van 13 vierkante centimeter.

figuur 4: open opgave uit het domein meten

Met behulp van de vragenlijst worden achtergrondkenmerken in kaart gebracht van de onderzoeksgroep van deelnemende leerkrachten en scholen.

Met betrekking tot het onderdeel leerlingkenmerken gaat het erom of de leerling allochtoon of autochtoon is, van welke sekse, om het aantal boeken thuis en of ze thuis de beschikking hebben over een computer. De vraag naar het aantal boeken is bedoeld als indicatie van het opleidingsniveau van de ouders, een kwestie die op een andere manier moeilijk te achterhalen is.

Uit eerdere TIMSS-studies is gebleken dat het achterhalen van het opleidingsniveau van de ouders bij leerlingen te veel ontbrekende waarden oplevert. Het aantal boeken dat de leerling thuis heeft, blijkt daar een redelijke indicatie voor te zijn.

Bij de achtergrondkenmerken van de leerkrachten gaat het om sekse, leeftijd, aantal jaren onderwijservaring en

specialisatie in de lerarenopleiding. Achtergrondkenmerken van de scholen zijn: grootte van de gemeente, economische achtergrond van de leerlingpopulatie, of het Nederlands als eerste taal van het gros der leerlingen moet worden beschouwd en de denominatie van de school, bijvoorbeeld 'rooms-katholiek'.

Instrumenten bij PISA

Met de PISA-toets wordt niet zozeer specifieke kennis getoetst, maar het vermogen om opgedane kennis en vaardigheden te gebruiken bij het oplossen van problemen. Om een indruk te krijgen van het soort opgaven dat in de PISA-toets zit, kunt u een selectie Nederlandse opgaven uit het hoofdonderzoek 2000 op de Nederlandse website² bekijken. Bijna de helft van de wiskundeopgaven in de toets bestaat uit meerkeuzevragen, de andere helft zijn open vragen en opdrachten. Ongeveer tweederde van de opgaven kan eenduidig 'goed' of 'fout' gescoord worden, voor ongeveer een derde deel van de opgaven en opdrachten is een correctievoorschrift met gedifferentieerde scoring opgesteld.

In figuur 5 volgt een voorbeeld - de opgave 'skateboard' - om een indruk te geven van het soort opdrachten.

Op de volgende pagina treft u in figuur 6 het scorevoorschrift aan.

SKATEBOARD		
Erik is een fanatiek skateboarder. Hij bezoekt een winkel met de naam 'SKATERS' om wat prijzen te bekijken. In deze winkel kun je een compleet skateboard kopen. Of je kunt een plank, een set van 4 wieltjes, een set van 2 wielassen en een set losse onderdelen kopen en je eigen skateboard monteren.		
De prijzen van de artikelen in de winkel zijn:		
Product	Prijs in zed	
compleet skateboard	82 of 84	
plank	40, 60 of 65	
een set van 4 wieltjes	14 of 36	
een set van 2 wielassen	16	
een set losse onderdelen (lagers, rubberen matjes, bouten en moeren)	10 of 20	

figuur 5: opgave 'skateboard' uit het PISA-onderzoek

De leerlingen krijgen bij de PISA-toets een vragenlijst voorgelegd met onder meer vragen over de leerstrategie die ze gebruiken bij het oplossen van de wiskunde problemen. Het gaat er daarbij om of de leerling onderscheid kan maken tussen hoofd- en bijzaken.

<p>Vraag 1: SKATEBOARD</p> <p>Erik wil zijn eigen skateboard monteren. Welke zijn de minimumprijs en de maximumprijs in deze winkel voor zelfgemonteerde skateboards?</p> <p>(a) Minimumprijs: ----- zed. (b) Maximumprijs: ----- zed.</p> <p>SKATEBOARD BEOORDELING 1</p> <p>Maximale score</p> <p>Code 21: Zowel het minimum (80) als het maximum (137) is juist.</p> <p>Gedeeltelijk goed</p> <p>Code 11: Alleen het minimum (80) is juist. Code 12: Alleen het maximum (137) is juist.</p> <p>Geen punten</p> <p>Code 00: Andere antwoorden. Code 99: Antwoord ontbreekt.</p>	<p>Vraag 2: SKATEBOARD</p> <p>De winkel verkoopt drie soorten planken, twee soorten sets met wieltjes en twee soorten sets losse onderdelen. Er is slechts één mogelijke keuze voor een set wielassen.</p> <p>Hoeveel verschillende skateboards kan Erik monteren?</p> <p>A 6 B 8 C 10 D 12</p> <p>SKATEBOARD BEOORDELING 2</p> <p>Maximale score</p> <p>Code 1: D. 12</p> <p>Geen punten</p> <p>Code 0: Andere antwoorden. Code 9: Antwoord ontbreekt.</p>
---	---

Vraag 3: SKATEBOARD

Erik mag 120 zed besteden en wil met dat geld het duurste skateboard kopen dat hij zich kan veroorloven.

Hoeveel geld kan Erik besteden aan elk van de 4 onderdelen? Noteer je antwoord in de onderstaande tabel.

Onderdeel	Bedrag (zed)
plank	
wiel­tjes	
wielassen	
losse onderdelen	

SKATEBOARD BEOORDELING 3

Code 1: 65 zed aan een plank, 14 aan wiel­tjes, 16 aan wielassen en 20 aan losse onderdelen.

Geen punten

Code 0: Andere antwoorden.
 Code 9: Antwoord ontbreekt.

figuur 6: eerste, tweede en derde vraag van de opgave ‘skateboard’, met scorevoorschrift

Voorts zijn er vragen die informatie moeten verschaffen over de houding van de leerling ten opzichte van het vak wiskunde. Bijvoorbeeld: Is wiskunde nuttig voor later? Vind je het leuk? Heb je zelfvertrouwen als je wiskunde-opgaven moet maken?

Overeenkomsten en verschillen

Bij alle toetsen vinden we open vragen en meerkeuzeopgaven. Bij PPON overheerst het aantal open opgaven. Bij PPON zijn de opgaven geordend in leerdomeinen, bij TIMSS in inhoudelijke en cognitieve categorieën en bij

PISA ligt het accent op het gebruik maken van kennis en vaardigheden.

In alle drie de onderzoeken wordt door middel van vragenlijsten aanvullende informatie verzameld. Bij PPON gaat het dan vooral om leerling- en schoolkenmerken, bij TIMSS eveneens, maar daar wordt ook informatie verzameld over leerkrachtkenmerken. Bij PISA wordt vooral informatie verzameld over de gevolgde leerstrategie en de houding van de leerling, dit alles met behulp van vragenlijsten die de leerlingen zelf moeten invullen.

3 De uitkomsten

De uitkomsten van PPON

De rapportage over de uitkomsten van de PPON-peiling hebben betrekking op het onderwijsaanbod, de resultaten per leerling, de verschillen tussen leerlingen en de effecten van methode en afnamejaar.

Het onderwijsaanbod

De invoering van de euro heeft geleid tot een snelle invoering van nieuwe of vernieuwde methoden. Daar was ook in 2003 al sprake van. Meer dan 80 procent van de scholen had op het moment van de peiling al een euromethode ingevoerd. Twee van de drie leraren in groep 3, 4 en 5 melden dat zij naast de reken-wiskundemethode nog andere leermiddelen gebruiken. De diversiteit daarvan is erg groot.

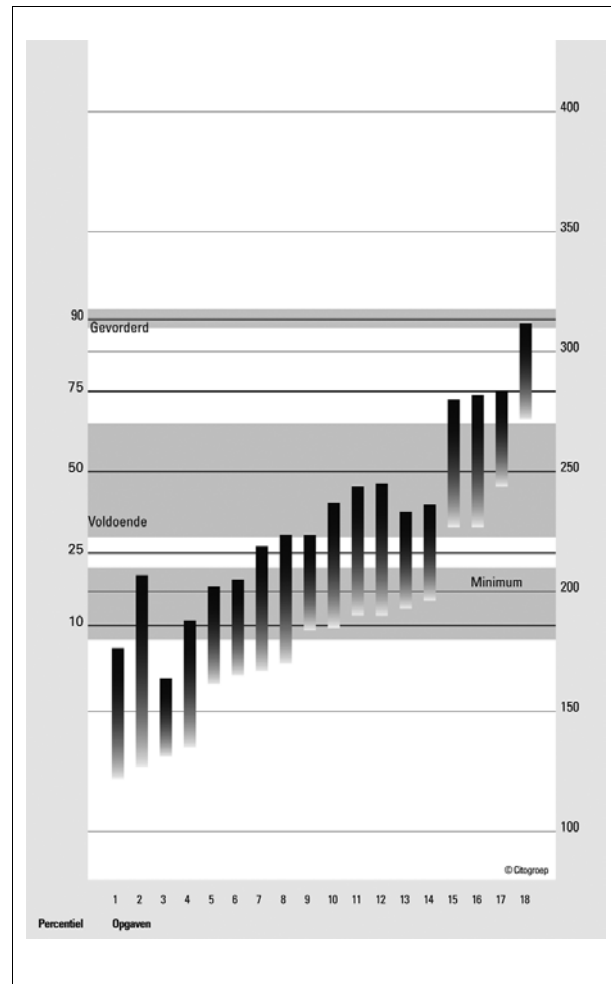
Interessant is de uitkomst met betrekking tot het aantal uren voor het vak rekenen-wiskunde. Vergeleken met 1987 is het aantal gemiddeld met een half uur toegenomen. Voor groep 3 betekent dit een stijging naar vier uur en voor de groepen 4 en 5 een stijging naar gemiddeld vierenhalf uur. Het is goed die toegenomen leertijd in gedachten te houden bij de beoordeling van de vergelijking tussen de diverse jaren. Er mag immers verwacht worden dat een toegenomen leertijd een stijging van de gemiddelde prestaties tot gevolg heeft.

Er is ook gekeken naar differentiatie. Opvallend is in de periode 1992-1997 de stijging van het percentage leraren dat kiest voor een organisatievorm waarbij de instructie in het algemeen voor alle leerlingen gelijk is, maar de verwerking verschilt naar niveau en tempo. Die stijging heeft zich voor de groepen 4 en 5 tussen 1997 en 2003 voortgezet. Dit is de dominante vorm van differentiëren, die wordt gepraktiseerd in circa 62 procent in groep 3 tot ruim 75 procent in groep 5.

Het pleidooi van deskundigen in het veld om minder aandacht te geven aan het traditionele cijferalgoritme en meer aan de inzichtelijke manier van het kolomsgewijs rekenen, lijkt voor wat betreft het aanbod bij optellen en aftrekken effect te hebben gehad. Dit effect zal voor het grootste deel (waarschijnlijk) toegeschreven moeten worden aan de invoering van die aanpak in de nieuwe of herziene methoden. Het is duidelijk dat het moment waarop met cijferen of kolomsgewijs rekenen wordt begonnen, verschoven is naar eind van groep 5 of groep 6. Bijna de helft (44 procent) van de leraren in groep 5 geeft aan dat aan beide vormen aandacht wordt gegeven. 32 procent zegt alleen het cijferalgoritme aan te leren en 8 procent zegt alleen de kolomsgewijze aanpak aan te bieden. 17 procent weet het niet, omdat deze algoritmen later aan de orde komen.

De resultaten per leerling

Voor elk van de onderwerpen die in paragraaf 2 zijn genoemd, worden de opbrengsten op verschillende manieren beschreven, beoordeeld en vergeleken op formatiegewicht, geslacht, de resultaten op de toetsen op vier tijdstippen van het leerlingvolgsysteem, en de afnamejaren van de PPON-peilingen. We nemen als voorbeeld de schaal 'bewerkingen': optellen en aftrekken tot 100 (fig.7). Rechts naast de grafiek staan de bekende standardscores met 250 als gemiddelde en verder met sprongen van 50 (een standaarddeviatie) naar boven en beneden. Links staan de percentielen voor 10, 25, 50, 75 en 90, aangegeven met een accentuering in een horizontale lijn in de grafiek.



figuur 7: de vaardigheidsschaal bij bewerkingen: optellen en aftrekken tot 100

Door opgaven uit het peilingsonderzoek met die van het proefonderzoek voor het leerlingvolgsysteem te combineren, zijn er voor het onderdeel bewerkingen optellen en aftrekken honderdtwaalf opgaven beschikbaar. De leerling met het percentiel 10 beheerst er daarvan dertig goed, de leerling met het percentiel 90 beheerst ze allemaal goed.

Met een goede beheersing wordt bedoeld dat die leerling

1
Opa had 68 euro in zijn portemonnee. Hij heeft voor 60 euro boodschappen gedaan.
Hoeveel euro heeft hij over?




2



Er zijn 36 verschillende plaatjes.
Nicky heeft al 25 plaatjes.
Hoeveel plaatjes mist zij nog?

3
De school gaat met 2 bussen op schoolreis.
In de ene bus zitten 50 leerlingen en in de andere bus 45.
Hoeveel leerlingen gaan mee?



figuur 8: voorbeeldopgaven 1, 2, 3 bij de vaardigheidsschaal optellen en aftrekken tot 100

9

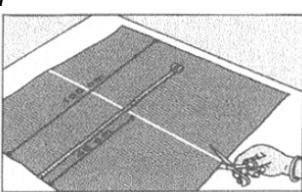


witte rozen rode rozen

Dit zijn bij elkaar 50 rozen. In de witte emmer staan 25 witte rozen.
Hoeveel rozen staan er dan in de grijze emmer?

rozen

17



De lap stof is 100 centimeter lang.
Moeder knipt een stuk stof af dat 48 centimeter lang is.
Hoe lang is het stuk dat ze overhoudt?

centimeter

18



De broek is goedkoper geworden.
Hoeveel euro is de broek goedkoper geworden?

euro

figuur 9: voorbeeldopgaven 9, 17, 18 bij de vaardigheidsschaal optellen en aftrekken tot 100

een kans van 80 procent of meer heeft om de opgave goed te beantwoorden.

Dat is iets anders dan de zekerheid dat die opgaven goed beantwoord worden. Het eenvoudigst is de wijze van rapporteren te begrijpen door te kijken naar de grafiek bij de vaardigheidsschaal bij een onderwerp. Wij nemen daarvoor optellen en aftrekken tot 100 (zie figuren 7, 8 en 9). Bij elke opgave hoort een kolommetje, donker aan de bovenkant en licht aan de onderkant. De bovenkant van de kolom geeft aan dat de kans op een goed antwoord 80 procent is. Bij opgave 17 betekent het dat een leerling met het percentiel 75 die opgave goed maakt. Dat wil zeggen de rekenvaardigheid van de leerling voor dit onderwerp zodanig is dat die een kans van 80 procent heeft om de opgave goed te beantwoorden. De onderkant van de kolom geeft aan een kans van 50 procent op een goed antwoord. Bij opgave 17 is die kans er voor een leerling met een percentielscore van 47.

Een prestatie wordt 'goed' genoemd als de kans op een goed antwoord 80 procent of hoger is. Ligt die kans tussen 50 en 80 dan wordt de prestatie matig genoemd. Is de kans op een goed antwoord kleiner dan 50 procent, dan wordt de prestatie getypeerd als onvoldoende.

Door te werken met kansen is het dus mogelijk dat een zwakke leerling toch een goed antwoord heeft - zij het dat

de kans daarop klein is - en dat een goede leerling dezelfde opgave fout beantwoordt (alweer is de kans daarop niet erg groot). Het is even wennen voor mensen die opgegroeid zijn met *p*-waarden (percentages goede antwoorden) gekoppeld aan opgaven. Daarin ligt het verschil. *P*-waarden zijn gekoppeld aan opgaven en kansredeneringen aan leerlingen of beter: aan het vaardigheidsniveau van leerlingen. Tot zover voor wat betreft de kolommen in de grafiek.

Rechts naast de grafiek staan de bekende standaardscores met 250 als gemiddelde en verder met sprongen van 50 (een standaarddeviatie) naar boven en beneden. Links staan de percentielen voor 10, 25, 50, 75 en 90, aangegeven met een accentuering in een horizontale lijn in de grafiek.

Daarmee zijn we er nog niet. In de grafiek zien we drie grijze banden met daarin de namen: minimum, voldoende en gevorderd. Dat zijn de zogenaamde standaarden.

De standaard 'voldoende' is de belangrijkste. Het is niet de verwachting dat alle leerlingen daaraan voldoen, maar het merendeel van de leerlingen dient dit niveau toch te bereiken. Bij de beschrijving is ervan uitgegaan dat dit niveau door 70 tot 75 procent van de leerlingen gehaald moet kunnen worden.

Het minimale niveau moet door vrijwel alle leerlingen

bereikt kunnen worden, dus door 90 tot 95 procent. Het niveau 'gevoerd' is van een andere orde. Voor die opgaven en de daarbij behorende onderwijshouden geldt dat die nog niet thuishoren in het leerplan voor midden groep vijf.

Gewapend met deze definities van voldoende, minimum en gevorderd heeft een panel van beoordelaars de standaarden vastgesteld. Het panel bestond uit veertien leraren met minimaal drie jaren ervaring in jaargroep 5, vijf schooladviseurs en zeven pabodocenten. Het eindoordeel, zoals weergegeven in de grafiek, kwam tot stand in een procedure van een eerste persoonlijke inschatting, daarop een inschatting na een discussie in kleine groepen en tot slot een oordeel na kennisname van de feitelijke toetsresultaten.

Die oordelen varieerden. De stroken geven het oordeel en de spreiding weer na weglating van de extremen. Het is duidelijk dat de feitelijke resultaten zijn achtergebleven bij de oordelen van het panel. Ofwel, de verwachtingen van het panel zijn veel te hoog. Pedagogisch gezien is het nuttig dat de verwachtingen hoger zijn dan de werkelijke resultaten, maar de discrepantie tussen verwachting en geconstateerd resultaat moet niet te groot zijn. Die standaarden vormen echter een risico wanneer we daaraan strikt de kwaliteit van het onderwijs afmeten.

We kijken nogmaals naar enkele opgaven, gewapend met bovenstaande kennis. Het balkje van opgave 2 in de vaardigheidsschaal is erg lang. Er is dus een aanzienlijke toename in vaardigheid nodig om van een 50 procent kans op een goed antwoord een kans van 80 procent te bereiken. Bij opgave 3 leidt een veel kleinere toename van de vaardigheid al tot een vergroting van de kans op een goed antwoord. Heeft dat te maken met de bekende buscontext? Of met de mooie getallen van opgave 3? Of met beide? We zien ook dat de opgaven uit deze grafiek tot het onderwijs van groep 5 behoren. De strook 'gevoerd' ligt boven het hoogste punt van het balkje van opgave 18. Er valt bij dit onderwerp verder nog te vermelden, en ook bij andere onderwerpen, dat leerlingen met een formatiegewicht 100 het beter doen, zoals te verwachten was, dan leerlingen met een formatiegewicht 190. Minder verwacht is de uitkomst van de vergelijking tussen de leerlingen met formatiegewicht 190 en 125. Daar zit hoegenaamd geen verschil tussen (formatiegewicht 190 iets beter, maar dat kan toeval zijn). Niet verrassend is eveneens dat jongens het beter doen dan meisjes. Dat is in overeenstemming met de toets als geheel. Bij deze schaal is er van 1992 tot 1997 geen vooruitgang - zelfs een lichte achteruitgang - te constateren, maar in 2003 is er weer sprake van vooruitgang ten opzichte van de vorige peilingen.

Verschillen tussen leerlingen

Het gaat hierbij om de bijdragen van verschillende variabelen aan de reken-wiskundige prestaties van leerlingen

die meegenomen zijn in de analyses van de gegevens. Gerapporteerd wordt in effectgrootten. Dat is zinvoller dan het melden of een verschil significant is. Met de grote aantallen observaties in PPOZ zijn verschillen immers al zeer snel significant.

Voor de kwalificatie van de effectgrootten geldt: 0,0 geen effect; 0,2 klein effect; 0,5 matig effect; 0,8 groot effect. Voor wie dat graag weten wil: de effectgrootte is het quotiënt van de gemiddelden en de standaardafwijking van de twee groepen die onderling vergeleken worden. Er zijn effectschattingen uitgevoerd voor zes variabelen: formatiegewicht, stratum, geslacht, leertijd, methode en afnamejaar. We gaan op elk van deze effecten in.

Formatiegewicht

In overeenstemming met de verwachting hebben leerlingen met de wegingsfactor 1,25 op alle onderwerpen een achterstand op 1.00-leerlingen met een gemiddelde effectgrootte van -0,47. De verschillen tussen 1.25-leerlingen en 1.90-leerlingen is bij de meeste onderwerpen verwaarloosbaar klein. Alleen bij het onderdeel verhoudingen is het verschil in het nadeel van de 1.90-leerling (-0,23). Het vrijwel ontbreken van verschillen tussen 1.25- en 1.90-leerlingen in de peiling van 2003 is opmerkelijk. In de vorige peiling was de achterstand van 1.90-leerlingen aanmerkelijk groter. Vermoedelijk kan dit gegeven verklaard worden uit een gecombineerd effect van het achteruit gaan van de 1.25-leerling en het vooruit gaan van de 1.90-leerling. De terugloop in prestaties van de 1.25-leerling is mogelijk het gevolg van de aanscherping van de definitie van 1.25-leerlingen. Was het eerst voldoende dat één ouder een lage vooropleiding had, nu geldt dat beide ouders een lage vooropleiding moeten hebben om een kind tot een 1.25-leerling te bestempelen.

Stratum

Scholen zijn ingedeeld in drie strata (zie hiervoor paragraaf 1) op basis van de schoolscores, die op hun beurt weer afgeleid zijn van leerlinggewichten. Welnu, het verschil tussen scholen met een verschillend stratumniveau voegt weinig of niets toe aan de verschillen die al verklaard worden door de verschillen tussen de leerlingen. Dat gegeven is belangwekkend. Als een school veel allochtone leerlingen telt, betekent dit niet een extra effect ten aanzien van de prestaties van de leerlingen. Die verschillen zijn, in tegenstelling tot wat veelal gedacht wordt, te verklaren uit het formatiegewicht van de afzonderlijke leerlingen.

Geslacht

Alleen voor het onderwerp 'basisautomatismen: vermenigvuldigen en delen' is er nauwelijks verschil tussen jongens en meisjes. Voor alle andere onderwerpen is er een significant verschil variërend van -0,09 tot -0,46, met een gemiddelde effectgrootte van -0,36. In 1997 was het verschil van vergelijkbare grootte. Er treedt in dit opzicht dus geen noemenswaardige verandering op.

Leertijd

Er is steeds weer een matig negatief effect voor de vertraagde leerlingen te constateren, vergeleken met hun jongere groepsgenoten.

Reken-wiskundemethode

Bij de analyse zijn alleen die methoden meegenomen waarvoor er in 2003 voldoende observaties waren. Die methoden zijn 'Pluspunt 1', 'Pluspunt 2', 'Rekenrijk', 'Rekenen en Wiskunde', 'Talrijk' en 'De wereld in getallen 1, 2 en 3'. Omdat de verschillen tussen 'Pluspunt 1 en 2' zeer klein bleken, zijn die twee versies in de vergelijking samengenomen. Hetzelfde geldt voor de methode 'De wereld in getallen 2 en 3'. De methoden zijn op acht onderwerpen met elkaar vergeleken waarbij de methode 'Pluspunt' als meest gebruikte methode op 0 is gesteld.

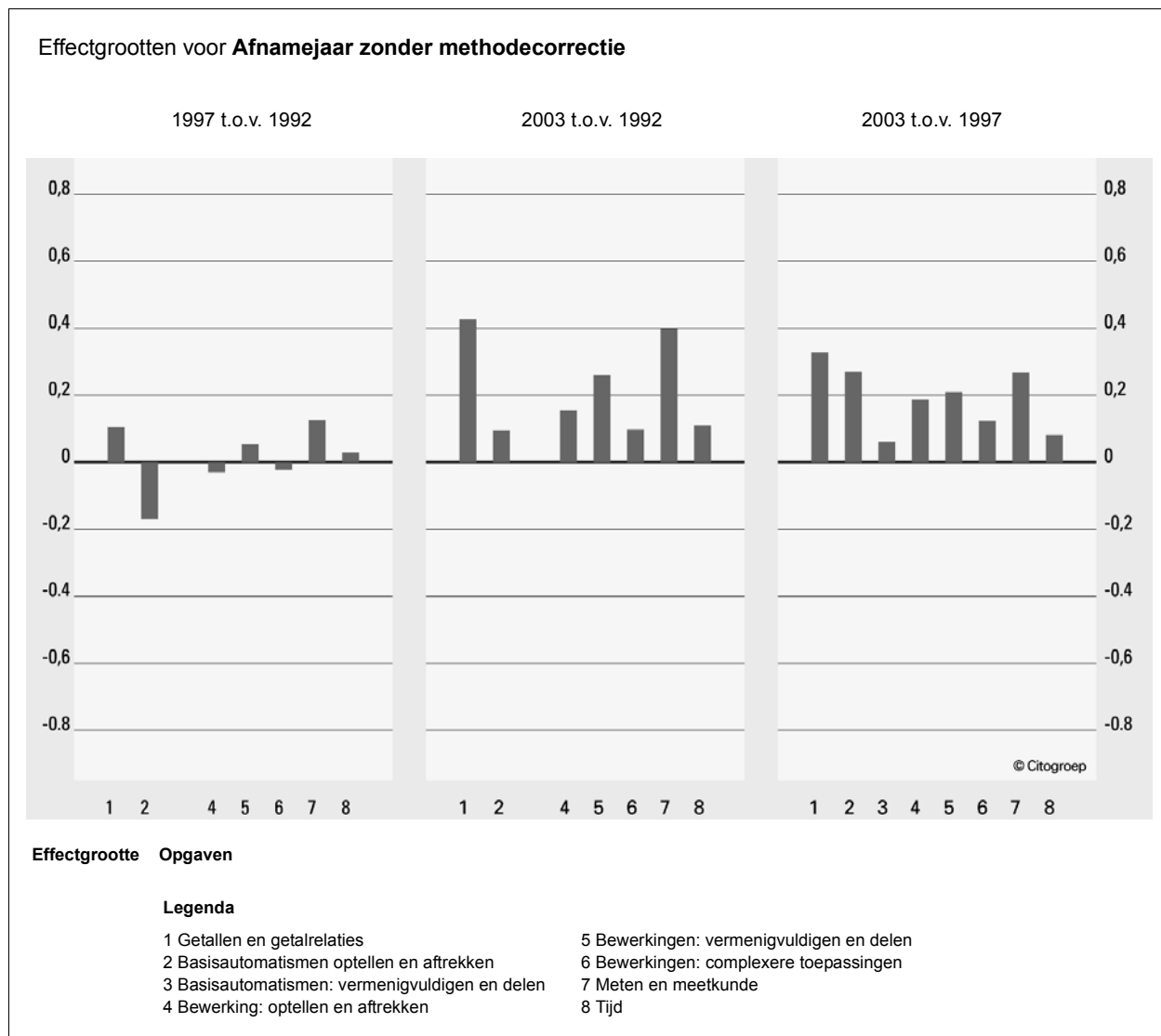
Bij deze vergelijking is er statistisch gecorrigeerd voor geslacht, leertijd, herkomst en afnamejaar. Dat is logisch,

want de methoden zijn niet gelijkmatig over scholen met verschillende leerlingbevolkingen verdeeld. De methode 'Rekenen en Wiskunde' komt relatief meer voor op scholen met veel 1.25- en 1.90-leerlingen. Dat effect moet dus gecorrigeerd worden.

De belangrijkste conclusie uit de vergelijkingen kan zijn dat het gebruik van meer recente methoden een positieve bijdrage levert aan de rekenvaardigheid van leerlingen. Bij de vorige peiling gold dat al voor de methoden 'Pluspunt' en 'De wereld in getallen'. Daar kunnen nu de methoden 'Talrijk' en 'Rekenrijk' aan worden toegevoegd.

Jaareffecten

Wanneer de effecten over de verschillende jaren in kaart worden gebracht, worden de uitkomsten daarvan statistisch gecorrigeerd voor de onderzochte variabelen die er invloed op hebben. Dat kan zeer terecht zijn als het bijvoorbeeld gaat om de herkomst van de leerlingen.



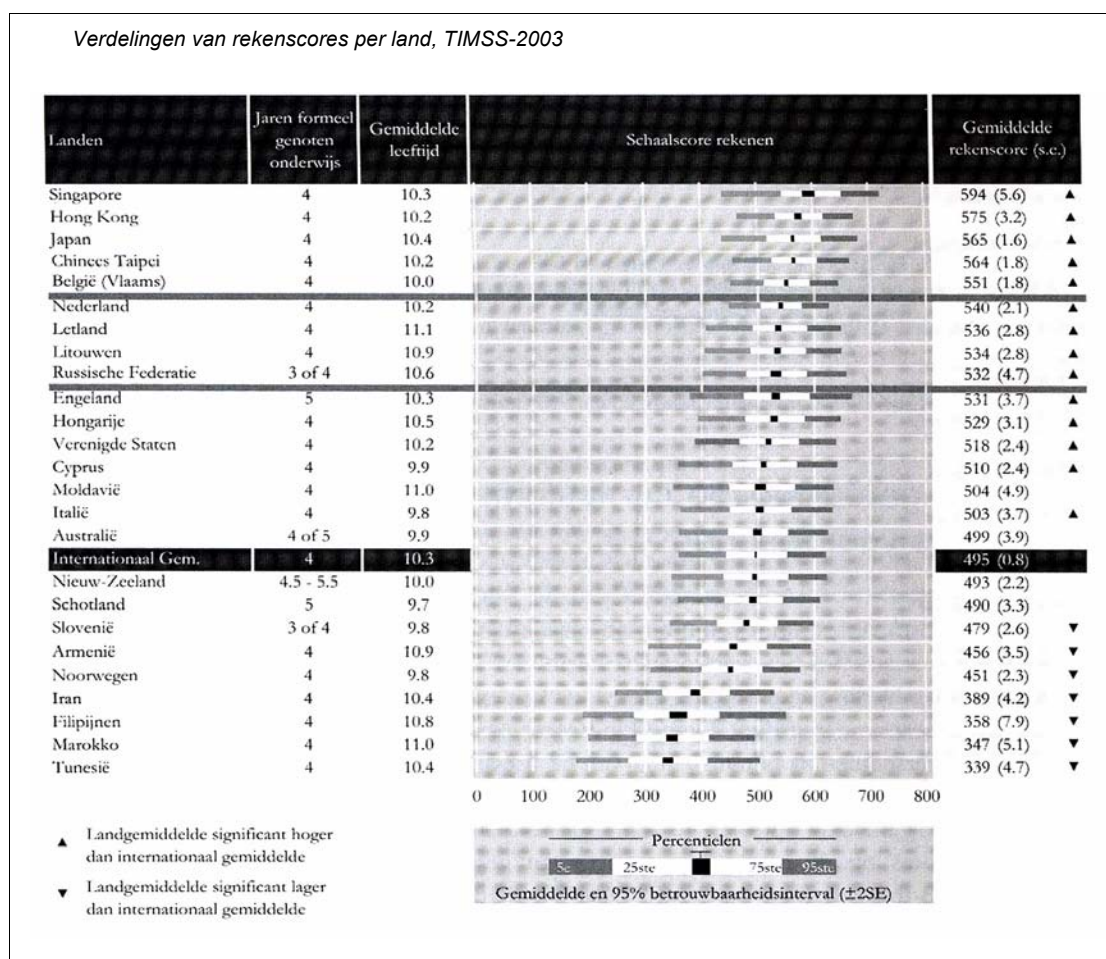
figuur 10: veranderingen in de prestaties volgens het PPON-onderzoek

Maar hoe staan we tegenover een correctie voor methode-effecten? Dankzij nieuwere methoden krijgen we betere resultaten voor het reken-wiskundeonderwijs, maar die halen we er van af om 'zuiver' te zijn. Wat is de redenering hier achter? De uitkomsten van het reken-wiskundeonderwijs worden door zeer veel zaken beïnvloed. Sommige daarvan zijn positief, andere negatief. Je kunt aan heel veel zaken denken. Om er enkele te noemen: de invloed van televisie, de kwaliteit van de lerarenopleiding, het aantal jaren ervaring in het onderwijs, of de school in de stad of op het platteland staat - we krijgen steeds meer stad en minder platteland - de hoeveelheid tijd die er aan rekenen wordt besteed, de interesse en aandacht van de ouders voor de leerresultaten van hun kinderen, de agenda van de kinderen - vaak zeer vol - enzovoort. De resultaten van het onderwijs zijn de resultante van deze en mogelijk nog vele andere factoren. In het PPOON-onderzoek is een beperkt aantal op voorhand relevant lijkende variabelen in het onderzoek en de analyse meegenomen. Over de niet meegenomen variabelen kunnen we op basis van dit onderzoek niets zeggen. Dat gebeurt dus ook niet. Maar als we de effecten van de wel geanalyseerde variabelen gebruiken om de uitkomsten daarvoor te corrigeren, krijgen we een beeld van hoe het reken-wiskundeonderwijs in de verschillende vergelij-

kingen er uit komt. Zo zijn de effectgrootten geschat in een vergelijking 1997 ten opzichte van 1992, die van 2003 ten opzichte van 1992 en die van 2003 ten opzichte van 1997. We zien dan dat de gemiddelde effecten overwegend zeer klein zijn. De combinatie van mogelijke positieve en negatieve invloeden heffen elkaar vrijwel op. Over de afzonderlijke invloeden is verder niets te zeggen. Maar er is alles voor te zeggen om de effecten van de methoden niet weg te zuiveren. De effecten van die methoden zijn een vaststaand feit. De meeste mensen zal het niet interesseren wat de uitkomst zou zijn als met dezelfde methoden als daarvoor was gewerkt. We mogen blij zijn met de positieve invloed van methoden. Er is zeer veel energie en kennis aan te pas gekomen om een ombuiging in de methoden te bewerkstelligen. Dat lijkt gelukt. Kijken we nu naar de jaareffecten zonder deze methodecorrectie, dan ziet het beeld er beter uit. Bij vergelijking van de uitkomsten van 2005 ten opzichte van 1992 en 1997 blijkt er wel degelijk sprake van vooruitgang (zie hiervoor fig.10).

Uitkomsten van het TIMSS-onderzoek

De rapportage over de uitkomsten van het TIMSS-onderzoek heeft betrekking op het onderwijsaanbod, resultaten



figuur 11: grafiek van pagina 24 scoreverdeling op de wiskundeschaal in de OESO- en partnerlanden, uit het TIMSS-onderzoek

die leerlingen van de basisschool boeken, op verschillen tussen leerlingen en de effecten van de methode en van het jaar van afname.

De leerlingresultaten

De toetsresultaten zijn weergegeven in de grafiek van figuur 11. Er wordt gerapporteerd in standaardscores. Het gemiddelde is gesteld op 500 met een standaarddeviatie van 100. De overeenkomst met de standaardscores van PPON valt op, maar in plaats van een gemiddelde van 250 met standaarddeviaties van 50 is bij TIMSS gekozen voor een gemiddelde van 500 met standaarddeviaties van 100. Die standaardscore vindt u onder de grafiek. In de grafiek staan balkjes voor ieder deelnemend land. Die balkjes geven het gemiddelde voor ieder land weer in zwart met de daarbij behorende standaardmeetfout. Voor Nederland is dat gemiddelde 540 met een standaarddeviatie van 2,1. Het betekent dat de werkelijke gemiddelde score van Nederland als alle leerlingen van groep 6 getoetst zouden zijn met een kans van 95 procent ligt tussen 535,89 en 544,2. Het gemiddelde + en - twee keer de standaarddeviatie. In de grafiek is te zien dat Nederland deel uitmaakt van een groep van vier landen waarvan de resultaten niet significant verschillen. In het balkje wordt niet alleen het gemiddelde aangegeven, maar ook de uitkomsten voor de percentiel -5, -25, -75 en -95 leerling. Opvallend bij Nederland is de lengte van het balkje. De spreiding is klein, kleiner dan bij de meeste andere landen. Zwakke rekenaars worden dus goed bij de les gehouden. Een goede zaak. Maar de goede leerling krijgt mogelijk onvoldoende uitdaging.

De uitkomsten van TIMSS 2003 zijn niet alleen vergeleken met die van andere landen, maar ook met de afname in 1995. Dan blijkt dat de toetsscore in groep 6 in 2003 significant is afgenomen. Het verschil is klein, van toetsscore 549 naar 540, maar die daling staat in contrast met de aanmerkelijk betere scores die een groot deel van de andere TIMSS-landen haalden.

Toetsprestaties van jongens en meisjes

In 1995 maakten de meisjes de rekentoets van TIMSS significant slechter dan jongens. Van de toen deelnemende landen waren de sekseverschillen in toetsprestaties in Nederland het grootst. In 2003 haalden de meisjes ook een significant lagere toetsscore, maar het verschil is veel kleiner geworden. Dat is niet te danken aan betere prestaties van de meisjes, maar aan een achteruitgang van de prestaties van de jongens. Meisjes presteerden in 2003 niet significant lager dan in 1995, jongens wel. Nederland is het enige land waar deze trendverschillen te zien zijn. Het kleiner geworden verschil tussen de prestaties van jongens en meisjes in Nederland heeft ertoe geleid dat in landen als de Verenigde Staten, Italië, Cyprus en Schotland dat verschil nu groter is dan in Nederland. Tegenover de landen waar de meisjes significant slechter presteerden dan jongens staan landen waar die situatie omgekeerd is. Bijvoorbeeld in Singapore, Filippijnen,

Moldavië en Armenië. Gemiddeld over alle 25 deelnemende landen presteren meisjes even goed als jongens. Kijken we naar de vijf inhoudelijke domeinen (zie paragraaf 2: getallen; patronen, vergelijkingen en verbanden; meten; meetkunde; gegevens), dan hebben de Nederlandse leerlingen de toetsonderdelen 'gegevens' en 'meten' het best gemaakt. Het minst goed werd op het onderdeel meetkunde gescoord. Vergeleken met meisjes blijken jongens significant beter te scoren in de domeinen 'getallen' en 'meten'.

Beoogd en uitgevoerd curriculum

In internationaal onderzoek is het moeilijk een toets te maken die spoort met de nagestreefde doelen van elke van de afzonderlijke landen. Daarom wordt nagegaan of onderwerpen en toetsopgaven passen binnen de nationale opvattingen. Twee curriculumexperts op het gebied van rekenonderwijs hebben onafhankelijk van elkaar de opgaven beoordeeld en hebben daarna overlegd over de opgaven waarover ze van mening verschilden. Het resultaat is een lijst met opgaven waarvan de experts van mening zijn dat de bijbehorende leerstof is onderwezen aan minimaal 50 procent van de leerlingen van groep zes. Volgens dat oordeel correspondeerden 26 van de 161 rekenopgaven niet met het nationaal beoogd curriculum. Daarmee past de TIMSS-toets van 2003 beter bij ons onderwijs dan de TIMSS-toets uit 1995.

Ook is door een groep leraren gekeken naar de toets en wel op twee manieren. Eerst is in een lijst van onderwerpen aangegeven wat behandeld is en daarna is een selectie van 31 rekenopgaven bekeken.

Uitkomsten van die beoordelingen:

- de toets van 2003 past beter bij ons onderwijs dan de toets van 1995;
- meetkunde, zoals ingevuld in de toets, wijkt nogal af van wat in ons land gangbaar is;
- de meetkundeopgaven worden minder geschikt bevonden; slechts een van de zes voorgelegde opgaven voldeed;
- de andere opgaven worden in voldoende mate geschikt bevonden.

Houding tegenover rekenen

Bij het TIMSS-onderzoek zijn vragen gesteld om na te gaan hoe de houding tegenover rekenen-meetkunde is. De internationale attitudevraag bestaat uit zes uitspraken, vier met betrekking tot het zelfvertrouwen en twee met betrekking tot plezier in rekenen.

Het blijkt dat leerlingen een licht positieve houding ten opzichte van rekenen hebben en veel zelfvertrouwen in eigen kunnen. Jongens tonen daarbij meer zelfvertrouwen dan meisjes. Jongens en allochtone leerlingen hebben iets meer plezier in rekenen dan meisjes en autochtone leerlingen. Allochtone leerlingen geven aan dat ze meer dan autochtone leerlingen aangemoedigd worden door hun leerkracht en door de ouders. Doordat

in het recente TIMSS-onderzoek andere vragen werden gesteld dan in 1995, is een vergelijking met de uitkomsten van 2003 niet mogelijk.

Uitkomsten van het PISA-onderzoek

Toetsresultaten

In de grafiek van figuur 12 zijn de landen gerangschikt volgens de score op *p*50, de gemiddelde score. Voor Nederland is af te lezen dat 90 procent (vanaf percentiel 5 tot en met percentiel 95) van de geteste leerlingpopulatie een score had tussen 385 en 683 met een gemiddelde van 538. Nederland staat daarmee op de vierde plaats. In de ons omringende landen blijft de score duidelijk achter. Maar worden voor België de scores van Vlaanderen en Wallonië apart bekeken, dan blijkt de Vlaamse score van 553 de hoogste van alle deelnemende landen. Vlaamse leerlingen blijken veel hoger te scoren dan Nederlandse.

Vaardigheidsniveaus

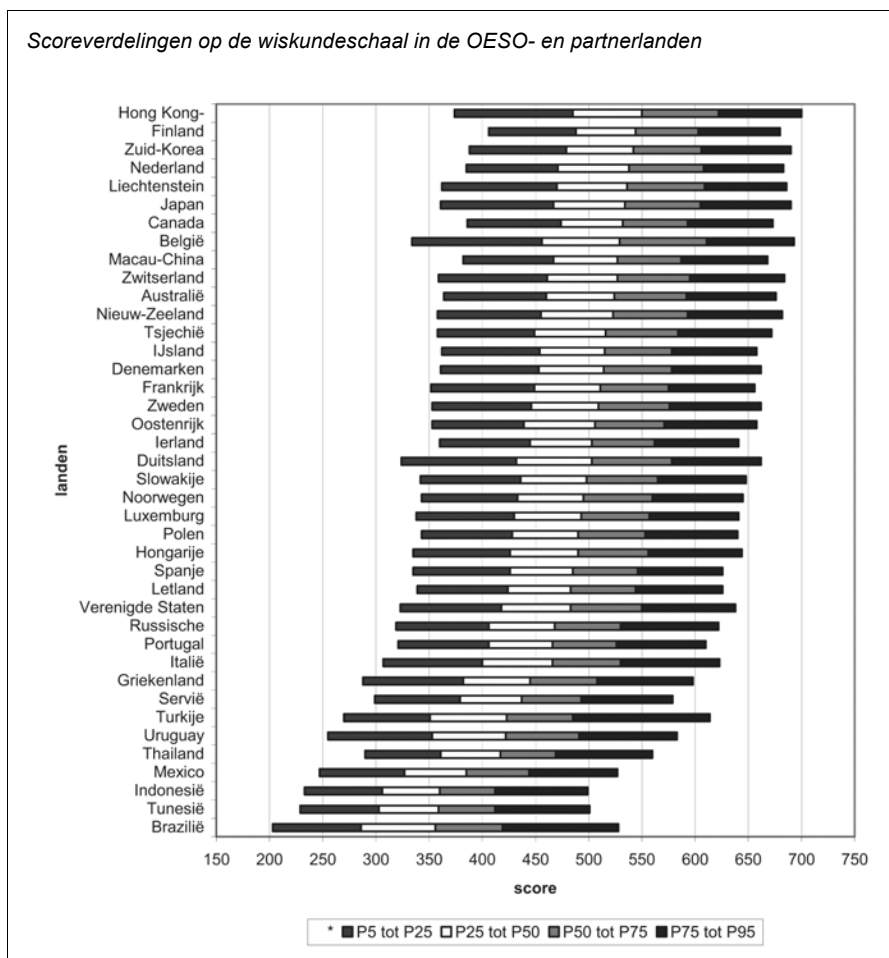
Voor de schaal uit de grafiek van figuur 12, over de scoreverdeling op de wiskundeschaal in verschillende landen, is een onderverdeling in zes vaardigheidsniveaus ontworpen. Niveau 1 van 385 tot 420, niveau 2 van 420 tot 482, enzovoort.

Die vaardigheidsniveaus zijn vervolgens getypeerd in beschrijvingen. Als voorbeeld geven we de beschrijving van schaal 4 (fig.13).

4	<p>Op niveau 4 kan een leerling:</p> <ul style="list-style-type: none"> - gericht werken met expliciete modellen van ingewikkelde situaties waarbij beperkingen aan de orde kunnen zijn of zelf veronderstellingen gemaakt dienen te worden; - kiezen uit, dan wel integreren van verschillende representatievormen, waaronder symbolische vormen, waarbij deze op een directe manier in verband gebracht kunnen worden met realistische situaties; - uitleg en argumenten construeren en communiceren, gebaseerd op eigen interpretatie en redeneringen. 	van 544 tot 606
---	--	-----------------

figuur 13: vaardigheidsniveau 4

Alle leerlingen worden daarna toebedeeld aan het hoogste niveau waarvan hij of zij naar alle waarschijnlijkheid het merendeel van de vragen correct beantwoordt. Wanneer de landen op basis van het percentage leerlingen in elk vaardigheidsniveau geordend worden,



figuur 12: internationale vergelijkingen in het PISA-onderzoek

dan blijkt Nederland weer hoog te scoren. Ons land komt op de vijfde plaats, na Finland, Korea, Canada en Hong Kong (China). In het rapport wordt gesteld dat in de Nederlandse samenleving mensen met een vaardigheidsniveau lager dan 4 niet voldoende toegerust zijn. Dat zou betekenen dat de helft van de Nederlandse leerlingen onvoldoende is toegerust voor onze Nederlandse samenleving, volgens de auteurs van dit rapport.

Het verschil tussen niveau 3 en 4 wordt gekarakteriseerd door een verschil in autonomie, in de mate van zelfstandigheid waarmee vraagstukken worden opgelost.

Resultaten per domein

In de toets zijn vier domeinen onderscheiden: (1) vorm en ruimte, (2) veranderingen en relaties, (3) onzekerheid en (4) hoeveelheid.

Bij het onderwerp 'vorm en ruimte' blijft Nederland op de ranglijst duidelijk achter met haar tiende plaats. Mogelijk wordt er in de onderbouw van het voortgezet onderwijs weinig aandacht aan dit onderdeel besteed. Voor het onderdeel 'veranderingen en relaties' staat Nederland daarentegen op de eerste plaats met een gemiddelde van 551. En bij het onderdeel 'onzekerheid' staat Nederland op de tweede plaats met alleen Hong Kong boven zich. Voor het onderdeel 'hoeveelheid' neemt Nederland de achtste plaats in. In het rapport wordt verder gekeken naar de resultaten per schooltype. Daar zijn geen verrassende resultaten te melden. Met het opleidingstype stijgt de gemiddelde score.

Enkele uitkomsten

We gaan nu in op enkele resultaten van het PISA-onderzoek, met name ten aanzien van het onderscheid tussen jongens en meisjes, de rol die taal speelt, de betekenis van het beroep en de herkomst van de ouders.

Jongens/meisjes

In PISA-2000 scoorden Nederlandse meisjes elf punten lager dan jongens. In PISA-2003 daalde dat verschil naar vijf punten, minder dan het PISA-gemiddelde van twaalf punten verschil. Dus, in vergelijking met de andere OESO-landen is het verschil in prestaties tussen jongens en meisjes in Nederland gering. Alleen in IJsland doen de meisjes het beter dan de jongens, met een verschil van vijftien punten.

Kijken we vervolgens naar de wiskundedomeinen, dan zien we dat de jongens in drie van de vier domeinen het beter doen dan de meisjes, maar in het domein 'hoeveelheid' is er een klein positief effect ten gunste van de meisjes. Alle verschillen zijn overigens klein.

Thuis gesproken taal

Er is voor wat de taal betreft in het onderzoek een onderscheid gemaakt tussen vier categorieën: Nederlands, een dialect, een Europese vreemde taal en een niet-Europese taal. De scores zijn respectievelijk 549, 551, 491, 464.

Het spreken van een dialect blijkt nauwelijks gevolgen te

hebben voor de prestaties. Het spreken van een vreemde taal daarentegen heeft wel effect. Dat effect is bovendien voor hen die een niet-Europese taal spreken het grootst. Opvallend, maar begrijpelijk is dit effect voor het domein 'hoeveelheid' het minste. Er wordt bij de opgaven uit dat domein immers minder een beroep gedaan op de taalvaardigheid van de leerlingen.

Beroep en opleiding ouders

Zoals te verwachten is vond men een duidelijke samenhang tussen beroep en opleidingsniveau van de ouders en de wiskundescores van de leerling. Hoe hoger het beroepsniveau van de ouders, des te hoger de wiskundescores van de leerlingen.

Vergelijking met de uitkomsten van PISA-2000

In 2000 zijn alleen de twee domeinen 'veranderingen en relaties' en 'vorm en ruimte' aan bod geweest. In 2000 scoorde Nederland het hoogst van alle deelnemende landen. Maar omdat het onderzoek van 2000 niet volledig aan de strenge eisen van de OESO voldeed, is dit resultaat niet in de officiële ranglijsten opgenomen. Er zijn analyses uitgevoerd om de resultaten van 2000 en 2003 toch vergelijkbaar te maken.

We zien dan dat Nederland, zowel in 2000 als in 2003, de hoogst gemiddelde score heeft op het domein 'veranderingen en relaties', maar die gemiddelde score van Nederland op dit domein is wel lager geworden. Van 568 naar 551 (standaardscores), een significante afname, terwijl elf andere landen in 2003 significant hoger scoorden dan in 2000. Dus een goed resultaat voor Nederland, maar zowel relatief als absoluut een achteruitgang ten opzichte van 2000. Voor het onderdeel 'vorm en ruimte' zakte Nederland van de vierde plaats op de ranglijst naar plaats zeven. Ook voor dit domein zien we zowel absoluut als relatief een lager resultaat in 2003.

Uit de vragenlijsten

In Nederland hebben jongens en meisjes, vergeleken met andere landen, een sterk uiteenlopend zelfbeeld over hun vaardigheid in wiskunde. Alleen in Liechtenstein, Zwitserland en Duitsland is dit in nog sterkere mate het geval. In Nederland verwacht minder dan 40 procent van de leerlingen dat het zich bezig houden met wiskunde een positieve invloed zal hebben op de rest van hun leven. Opvallend verschijnsel daarbij is dat meisjes die het zeer eens zijn met positieve uitspraken over het belang van wiskunde gemiddeld slechtere resultaten hebben dan meisjes die het er mee eens zijn (dus niet zeer mee eens). Blijkbaar vinden meisjes naarmate ze meer weten van wiskunde dat die wiskunde niet van belang is voor het krijgen van een baan.

Ander opmerkelijk feit is dat in landen waar leerlingen ondergebracht zijn in verschillende schooltypes leerlingen in op de arbeidsmarkt georiënteerde scholen zich minder thuis voelen op school dan leerlingen in het algemeen vormend onderwijs. In Nederland, Finland en

Zwitserland hebben meisjes weinig zelfvertrouwen als het gaat om wiskunde en Nederlandse leerlingen staan negatief ten opzichte van de hoeveelheid steun die ze van de leraar krijgen. Opvallend is dat leerlingen die melden dat de leraar elke les interesse toont, in de voortgang van de leerlingen lager scoren dan leerlingen die aangeven dat dit in de meeste of sommige lessen of nooit gebeurt.

4 Slotopmerkingen

In de beide internationale vergelijkende onderzoeken, PISA en TIMSS, blijkt Nederland goed te scoren. Wel is er, vergeleken met vorige peilingen, sprake van een lichte daling.

Bij de PPON-medio-toets daarentegen is er, vergeleken met eerdere peilingen, sprake van een stijging. Is dit het gevolg van dat wat hierin getoetst wordt beter overeenkomt met wat wordt onderwezen? Bij internationale onderzoeken is die overeenkomst uiteraard minder.

Verwachtingen

Zowel bij het PPON- als bij het OESO-onderzoek zijn de verwachtingen van beoordelaars hoger dan de feitelijke uitkomsten. Dit verschijnsel doet zich steeds weer voor. De verwachtingen zijn mogelijk wat overspannen.

Jongens/meisjes

Bij het PPON-onderzoek is er geen noemenswaardige verandering als het gaat om het verschil in prestaties tussen jongens en meisjes, bij het TIMSS-onderzoek was het verschil in prestaties tussen jongens en meisjes in Nederland in 1995 het grootst van alle deelnemende landen.

Dat verschil is bij het onderzoek van 2003 veel kleiner geworden, hetgeen voor Nederland vooral te wijten is aan de dalende prestaties van de jongens. Ook bij het PISA-onderzoek was sprake van een afnemend verschil.

Noten

- 1 Zie ook: www.timss.com
- 2 Zie ook: www.pisa.nl

Literatuur

- Gille, E., K. Lagerwaard, G. Limpens, A. de Knecht-van Eekelen, G. Maris & C. van Rhijn (samenstelling) (2004). *Resultaten PISA-2003, praktische kennis en vaardigheden van 15-jarigen. Nederlandse uitkomsten van OESO Programme for International Student Assessment (PISA) op het gebied van wiskunde, leesvaardigheid, natuurwetenschappen en probleem oplossen in het jaar 2003*. Arnhem: Cito-groep.
- Meelissen, M.R.M. & B.G. Doornekamp (2004). *TIMSS-2003 Nederland. Leerprestaties in exacte vakken in het basisonderwijs*. Enschede: Universiteit Twente.
- Kraemer, J-M., J. Jansen, F. van der Schoot & B. Hemker (2005). *Balans (31) van het reken-wiskundeonderwijs halverwege de basisschool 4. PPON-reeks 31*. Arnhem: Cito-groep.

In this contribution three tests are compared, all of them targeting the level achieved by students in mathematics education. PPON is a Dutch test, which is held every five years. Its goal is to compare the results of mathematics education in primary schools (grade 5 and 8; age about ten and twelve) over time. TIMSS and PISA are aimed at comparing the results of mathematics teaching across several countries. TIMSS focuses on primary education, PISA on secondary education. Compared to their peers in other countries - both Western and non-Western - Dutch children seem to achieve better results than many others. The good results of Dutch children in national and international tests in mathematics can probably be connected to the specific didactical approach in Dutch mathematics textbooks in primary and secondary schools: realistic mathematics education.