

Enige tijd geleden zijn de uitkomsten van de internationale vergelijkende studie TIMSS bekend gemaakt in de landelijke pers. Nederland bleek het goed gedaan te hebben in vergelijking met andere landen. **Wilma Kuiper, Klaas Bos en Tjeerd Plomp** gaan in op de 'nationale optietoets', een aanvullende toets die alleen in Nederland in het tweede leerjaar van het voortgezet onderwijs is afgenomen.

De TIMSS nationale optietoets wiskunde

Wat is TIMSS?

De *Third International Mathematics and Science Study* (TIMSS) is een internationaal vergelijkend onderzoek naar de opbrengst van het onderwijs in wiskunde en de natuurwetenschappelijke vakken. Aan TIMSS, dat wordt uitgevoerd onder auspiciën van de *International Association for the Evaluation of Educational Achievement* (IEA), nemen ruim veertig landen deel. Het Nederlandse aandeel in de studie wordt uitgevoerd door het Onderzoek Centrum Toegepaste Onderwijskunde (OCTO) van de Universiteit Twente en richt zich onder meer op het eerste en tweede leerjaar voortgezet onderwijs. Binnen TIMSS wordt dit onderdeel aangeduid als *populatie 2*.

Het hoofdonderzoek heeft plaatsgevonden in het voorjaar van 1995. In Nederland waren daar 95 scholen voor voortgezet onderwijs bij betrokken. De gegevensverzameling bestond onder meer uit de afname van een internationale wiskunde- en sciencetoets in één eerste en één tweede klas per school, van internationale vragenlijsten voor leerlingen, leraren en schoolleiders, en van een internationale praktische vaardigheidstoets (in 49 tweede klassen).

De internationale wiskunde- en sciencetoets bestond uit 150 wiskunde-opgaven en 135 science-opgaven, verdeeld over acht toetsboekjes. Een leerling kreeg één toetsboekje voorgelegd. Elk boekje bevatte een deel van zowel de wiskunde- als science-opgaven. Een bepaalde rotatiesystematiek waarborgde de onderlinge vergelijkbaarheid van de verschillende boekjes. De opgaven bestonden uit meerkeuze- en open vragen. De 150 wiskunde-opgaven, verder aangeduid als de internationale wiskundetoets, dekten zes leerstofonderdelen (breuken en getalbegrip; meetkunde; algebra; datarepresentatie, data-analyse en waarschijnlijkheid; meten; verhoudingen) en vier categorieën van vaardigheden (kennis en inzicht; uitvoeren van routine-matige handelingen; uitvoeren van complexe handelingen; onderzoeken en probleemoplossen). Het internationale rapport over TIMSS-populatie 2 is eind 1996 uitgebracht (Beaton e.a., 1996), het Nederlandse rapport in juni 1997 (Kuiper, Bos & Plomp, 1997).

Waarom een nationale optietoets?

Medio 1994, in de aanloop tot vaststelling van de internationale wiskundetoets voor populatie 2, trok de begeleidingscommissie van het Nederlandse TIMSS-project de conclusie dat de geschiktheid van de proefversie van de toets voor Nederland veel te wensen overliet. Het nieuwe wiskundeprogramma voor de onderbouw van het voortgezet onderwijs werd, zo oordeelde de commissie (met onder andere vertegenwoordigers van het Freudenthal Instituut en het CITO), onvoldoende door de internationale proeftoets gedekt.

De kritiek richtte zich op zowel de *vorm* als de *inhoud*. De toets zou te veel het accent leggen op het reproduceren van feitenkennis, het kunnen uitvoeren van meer of minder complexe routineprocedures en het oplossen van relatief eenvoudige wiskundige problemen en te weinig op redeneren en bewijzen (*inhoud*). Bovendien zou bij het nieuwe programma geen toetsing passen in de vorm van meerkeuze-opgaven en werd de kwaliteit van een aantal korte open vragen in twijfel getrokken (*vorm*).

Vanwege de *geringe representativiteit* van de internationale wiskundetoets voor Nederland vreesde de commissie dat Nederlandse leerlingen lager op de toets zouden presteren dan op grond van hun niveau kon worden verwacht. Besloten werd een nationale optietoets wiskunde voor leerjaar 2 te ontwikkelen en als negende toetsboekje toe te voegen aan de internationale wiskundetoets. De nationale optietoets diende te bestaan uit wiskunde-opgaven die representatief zijn voor het nieuwe Nederlandse wiskundeprogramma.

De constructie van de nationale optietoets is in november 1994 in een samenwerkingsverband van Freudenthal Instituut, CITO, APS en OCTO ter hand genomen. De uiteindelijke versie van de toets bestond uit zestien zogenaamde *ankeropgaven* (wiskunde-opgaven uit de internationale wiskundetoets die passen bij het nieuwe wiskundeprogramma) en twintig *nationale opgaven* (speciaal voor dit onderzoek ontwikkelde en beproefde wiskunde-opgaven, die eveneens dekkend zijn voor het nieuwe wiskundeprogramma). De toets is in het voorjaar van 1995 tege-

lijk met de internationale wiskundetoets afgenomen in alleen het tweede leerjaar. Per klas kreeg elke leerling aselekt één van de negen toetsboekjes toegewezen.

Vraagstelling

Gegeven de twijfel over de representativiteit van de internationale wiskundetoets voor Nederland en, belangrijker nog, de vrees voor onderschatting van de wiskundeprestaties van Nederlandse leerlingen op basis van de internationale toets, had de afname van de nationale optietoets wiskunde tot doel een antwoord te formuleren op de volgende twee vragen:

1. Hoe scoren Nederlandse leerlingen in het tweede leerjaar voortgezet onderwijs in vergelijking met leerlingen in enkele andere landen op de ankeropgaven?
2. In hoeverre treden er verschillen op in prestaties van Nederlandse leerlingen in het tweede leerjaar voortgezet onderwijs op de nationale opgaven respectievelijk internationale wiskundetoets?

Gebruik is gemaakt van een zogenaamd ankeropgaven-design (Tabel 1; zie Eggen, 1993). Doel van het anker is een vergelijking mogelijk te maken tussen de prestaties van Nederlandse tweedeklassers op de nationale opgaven respectievelijk de internationale wiskundetoets.

De lichtgrijs gearceerde cellen 2, 5 en 8 hebben betrekking op de eerste vraag. Het gaat hier om een vergelijking van de prestaties van Nederlandse tweedeklassers (cel 2

en 5) met die van een vergelijkbare groep leerlingen uit enkele andere landen (cel 8). Omdat het ondoenlijk was alle TIMSS-landen in deze vergelijking te betrekken, is om praktische redenen gekozen voor drie buurlanden: Vlaanderen, Engeland en Denemarken. De Nederlandse leerlingen in cel 5 en de Vlaamse, Engelse en Deense leerlingen in cel 8 hebben de ankeropgaven gemaakt als onderdeel van de internationale wiskundetoets; de Nederlandse leerlingen in cel 2 hebben de ankeropgaven gemaakt als onderdeel van de nationale optietoets. Verondersteld wordt dat er geen verschillen in prestaties optreden tussen de leerlingen in cel 2 en cel 5. Het gaat immers om dezelfde zestien opgaven en om een naar samenstelling vergelijkbare groep leerlingen.

Een tweede veronderstelling is dat de verschillen in prestaties op het anker tussen de cel 5 en cel 8 leerlingen ongeveer in dezelfde orde van grootte zullen liggen als de verschillen in prestaties op de internationale toets tussen de cel 6 en cel 9 leerlingen.

Deze twee veronderstellingen vormen de basis voor een vergelijking van de prestaties van Nederlandse leerlingen op de nationale opgaven (cel 1) met die van Nederlandse leerlingen op de internationale wiskundetoets (cel 6). Juist om deze vergelijking draait het allemaal.

De donkergrijs gearceerde cellen 1 en 6 verwijzen naar de tweede vraag. Verondersteld wordt dat de leerlingen die de nationale opgaven hebben gemaakt – juist omdat die opgaven representatief zijn voor het nieuwe wiskundeprogramma – beter in internationale context zullen

<p><i>cel 1</i></p> <p>20 nationale opgaven, als onderdeel van de nationale optietoets (boekje 9); gemaakt door Nederlandse tweedeklassers (n = 239)</p>	<p><i>cel 2</i></p> <p>16 ankeropgaven, als onderdeel van de nationale optietoets (boekje 9); gemaakt door Nederlandse tweedeklassers (n = 239)</p>	
	<p><i>cel 5</i></p> <p>16 ankeropgaven, als onderdeel van de internationale wiskundetoets (boekje 1-8); gemaakt door Nederlandse tweedeklassers (n = 1954)</p>	<p><i>cel 6</i></p> <p>internationale wiskundetoets (boekje 1-8); gemaakt door Nederlandse tweedeklassers (n = 1954)</p>
	<p><i>cel 8</i></p> <p>16 ankeropgaven, als onderdeel van de internationale wiskundetoets (boekje 1-8); gemaakt door <i>upper grade</i> leerlingen in Vlaanderen (n = 2894), Engeland (n = 1777) en Denemarken (n = 2344)</p>	<p><i>cel 9</i></p> <p>internationale wiskundetoets (boekje 1-8), gemaakt door <i>upper grade</i> leerlingen in andere landen, waaronder Vlaanderen, Engeland en Denemarken</p>

presteren dan de leerlingen die de voor Nederland als minder geschikt aangemerkte internationale wiskunde-toets hebben gemaakt. Met 'in internationale context' wordt bedoeld: indien leerlingen in andere landen deze nationale opgaven ook zouden hebben gemaakt.

De vergelijking tussen de cellen 1 en 6 respectievelijk 9 is echter niet direct te maken, aangezien de nationale opgaven niet door leerlingen in andere landen zijn gemaakt. De vergelijking loopt via het anker, dus via de cellen 2 en 5 respectievelijk 8.

De leerlingen in de cellen 1 en 2 zijn de leerlingen die de nationale optietoets hebben gemaakt. De vergelijking van prestaties van Nederlandse leerlingen uit leerjaar 2 op de internationale wiskundetoets (cel 6) met die van leerlingen uit de *upper grade* in andere landen (cel 9) is onderdeel van het reguliere Nederlandse aandeel in TIMSS. De resultaten ervan staan uitgebreid beschreven in het onlangs verschenen nationale rapport (Kuiper, Bos & Plomp, 1997). Hier beperken we ons tot de algemene conclusie dat Nederlandse tweedeklassers, net als de eersteklassers, het goed doen op de internationale wiskundetoets.

De nationale opgaven

Het *Trajectenboek Wiskunde 12-16* (Commissie-Ontwikkeling Wiskundeonderwijs, 1992) vormde de inhoudelijke basis voor de *nationale opgaven*. De opgaven zijn verder gebaseerd op een onderscheid in drie beheersingsniveaus (De Lange & Boertien, 1994; De Lange, 1995):

- een laag niveau dat zich voornamelijk beperkt tot reproductief handelen en het toepassen van eenvoudige berekeningen en algoritmen (bijvoorbeeld geven van een definitie, uitvoeren van eenvoudige berekeningen, tekenen van eenvoudige grafieken)
- een midden niveau, waarop eenvoudige 'echte' problemen worden opgelost waarbij leerlingen zelf enige verbanden moeten leggen of zelf keuzen moeten maken
- een hoog niveau, waarbij het gaat om procesvaardigheden als mathematiseren dan wel vertalen van een redelijk complex probleem naar een wiskundige kern, redeneren, communiceren (verwoorden van een oplossing), integreren van verschillende vakonderdelen, reflecteren enzovoort.

Bij de constructie van de opgaven is ten aanzien van de beheersingsniveaus rekening gehouden met grote verschillen in de beheersing van basiskennis en basisvaardigheden per schooltypegroep. Gekozen is voor een relatief groot aantal opgaven op laag niveau (elf), een kleiner aantal opgaven op het midden niveau (zes) en een nog kleiner aantal opgaven op het hoge niveau (drie). Figuur 1 visualiseert de verdeling van de twintig nationale opgaven over de drie beheersingsniveaus en de vier in het Trajectenboek onderscheiden leerstofonderdelen.

De twintig nationale opgaven bestaan uitsluitend uit open vragen en zijn gegroepeerd rond zes contexten (tussen haakjes staan de opgavennummers vermeld (zie figuur 1):

- Mozaïekstenen: vier opgaven (17-20), waarvan één over meetkunde (17) en drie op het snijvlak van algebra en meetkunde (18-20)
- Tijdverschil: drie opgaven over rekenen (1-3)
- Kandelaar: drie opgaven over kijkmeetkunde (4-6)
- Lengte: drie algebra-opgaven (7-9)
- Juli 1994: vier opgaven over statistiek (10-13)
- Fietstocht: drie opgaven, waarvan één over algebra en het aflezen van grafieken (14) en twee op het snijvlak van rekenen en algebra (15-16).

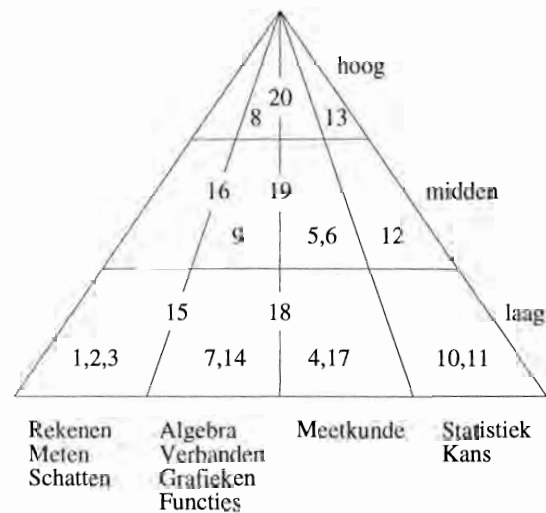


fig. 1 Beheersingsniveaus en leerstofonderdelen: verdeling van de nationale opgaven uit de nationale optietoets wiskunde

Een proefversie van de nationale optietoets is in januari 1995 afgenomen bij 271 leerlingen, verdeeld over drie scholen en elf tweede klassen van verschillende samenstelling (van VBO/MAVO, MAVO, HAVO en VWO). Om een indruk te krijgen van het soort nationale opgaven staan hieronder de opgaven Mozaïekstenen en Lengte weergegeven.

Mozaïekstenen

Hieronder zie je een tekening van een vierkante mozaïeksteen met een rond gat in het midden.



17. Kun je met precies 65 van deze stenen een rechthoek leggen?
Leg uit waarom wel of waarom niet.
18. Kun je met precies 65 van deze stenen een vierkant leggen?
Leg uit waarom wel of waarom niet.

Er zijn ook mozaïekstenen met meer gaten erin. Hieronder zie je een tekening van zo'n steen.



19. Met 9 van deze mozaïekstenen wordt een vierkant gelegd.
Hoeveel hele cirkels kun je dan zien?
Laat zien hoe je aan je antwoord komt.
20. Hoeveel cirkels kun je zien als je een vierkant maakt van 10 bij 10 stenen?
Leg uit hoe je aan je antwoord komt.

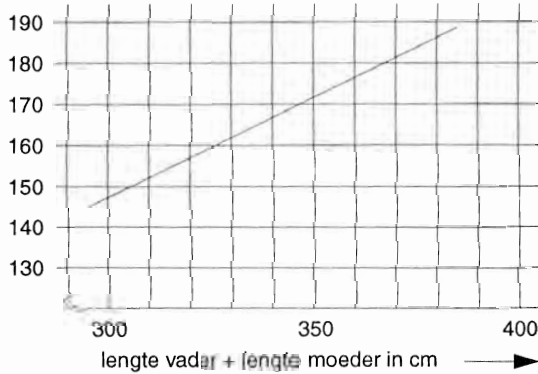
Lengte

Om te berekenen hoe lang een meisje ongeveer zal worden, gebruikt een schoolarts de volgende formule:

$$\begin{aligned} \text{lengte dochter (in cm)} &= \\ &= \frac{\text{lengte vader (cm)} + \text{lengte moeder (cm)} - 12}{2} + 3 \end{aligned}$$

7. De vader van Daniëlle is 1,82 m lang, haar moeder is 1,68 m. Hoe lang zal Daniëlle volgens deze formule worden?
8. Is het mogelijk dat – volgens de formule – een dochter groter wordt dan haar vader?
Licht je antwoord toe. Dat mag ook met een voorbeeld.

Hieronder is de grafiek getekend die hoort bij de formule.



Wanneer de dochter kleiner blijft dan 1,57 m is ze 'nogal klein'. Wanneer de dochter groter wordt dan 1,80 m is ze 'nogal groot'.

9. Geef in de tekening hierboven op de grafiek aan welk gedeelte hoort bij 'nogal klein' en welk gedeelte hoort bij 'nogal groot'.
Gebruik een kleurpotlood en zet erbij 'nogal klein' en 'nogal groot'.

De ankeropgaven

De belangrijkste eisen bij de samenstelling van het anker waren (1) dat het diende te bestaan uit internationale opgaven die inhoudelijk zo goed mogelijk pasten bij het nieuwe wiskundeprogramma en (2) dat de geselecteerde opgaven zoveel mogelijk gespreid waren over de leerstofonderdelen en deelgebieden die in de internationale wiskunde-toets worden onderscheiden.

Het uiteindelijke anker bestond uit zeven opgaven over breuken en getalbegrip, twee over meetkunde, drie over algebra, twee over datarepresentatie, -analyse en waarschijnlijkheid, één over meten en één over verhoudingen. Twaalf opgaven waren meerkeuzevragen, vier bevatten open vragen (drie kort-open en één lang-open). Vier ankeropgaven staan bij wijze van voorbeeld hieronder weergegeven.

Opgave O09

Gewone breuken: operaties, relaties en eigenschappen
Probleemoplossen

Peter traint elke dag door 5 km hard te lopen. De baan waarop hij loopt is $\frac{1}{4}$ km lang.

Hoeveel rondjes loopt hij elke dag op de baan?

Opgave V02

Datarepresentatie en -analyse
Redeneren

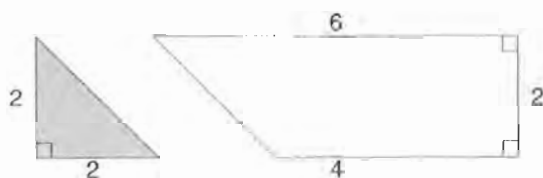
Deze twee advertenties stonden in de krant van een land waar de zed de munteenheid is.

Gebouw A	Gebouw B
Kantoorruimte te huur	Kantoorruimte te huur
Als u kantoorruimte tussen 85-95 vierkante meter huurt, dan kost u dat 475 zeds per maand	Als u kantoorruimte tussen 35-260 vierkante meter huurt, dan kost u dat 90 zeds per vierkante meter per jaar
Als u tussen 100-120 vierkante meter huurt, dan kost u dat 800 zeds per maand	

Een bedrijf in dat land wil een kantoor van 110 vierkante meter voor een jaar huren.

In welk gebouw, A of B, zouden ze de kantoorruimte moeten huren om het goedkoopst uit te zijn?
Laat zien hoe je aan je antwoord komt.

Opgave R10
Overige meetkunde
Complexe handelingen



Hierboven is een grijsgemaakte driehoek getekend en daarnaast een trapezium. Hoeveel van deze driehoeken passen precies in het trapezium?

- A. Drie
- B. Vier
- C. Vijf
- D. Zes

Opgave K06
Schatten en getalbegrip
Complexe handelingen

Vorig jaar bedroeg het aantal leerlingen op het Thorbecke College 1172. Dit jaar is het aantal leerlingen ongeveer 15 procent hoger dan vorig jaar. Hoeveel leerlingen heeft het Thorbecke College dit jaar ongeveer?

- A. 1800
- B. 1600
- C. 1500
- D. 1400
- E. 1200

Opzet van het onderzoek

De nationale optietoets is in april/mei 1995 als negende toetsboekje afgenomen in de tweede klassen waarin tegelijkertijd de internationale wiskundetoets is afgenomen. Per klas kregen twee à vier leerlingen (elke negende leerling) de nationale optietoets voorgelegd. Iedere overige leerling uit die klas maakte één van de acht internationale boekjes. De nationale optietoets is gemaakt door 239 leerlingen (zie cel 1 en 2 in tabel 1). Op basis van zijn of haar schooltype-advies is een leerling toegewezen aan de schooltypegroep VBO/MAVO of HAVO/VWO. Van de 239 leerlingen behoren er 140 tot VBO/MAVO-2 en 99 tot HAVO/VWO-2.

De afname van de nationale optietoets nam, net als bij de internationale toets, twee aaneengesloten lesuren in beslag. De nationale optietoets bestond, net als de internationale toets, uit twee delen, met daar tussenin een korte pauze. Het eerste deel van de nationale optietoets bevatte

de ankeropgaven plus vier nationale opgaven. Het tweede deel bevatte de overige zestien nationale opgaven. Het gebruik van een zakrekenmachine, geodriehoek, passer en dergelijke was niet toegestaan bij het eerste deel van de toets, maar wel bij het tweede deel. Bij het tweede deel werd dat toegestaan omdat het gebruik van hulpmiddelen en materialen onlosmakelijk deel uitmaakt van het nieuwe wiskundeprogramma. Bij het eerste deel mocht dat niet, aangezien dat ook niet was toegestaan aan leerlingen die de internationale toets maakten. Uniforme afname was noodzakelijk met het oog op vergelijkbaarheid van prestaties.

De eerste onderzoeksvraag

Bij de eerste onderzoeksvraag gaat het, zoals gezegd, om een vergelijking van de prestaties van drie groepen leerlingen op de ankeropgaven: Nederlandse leerlingen die de ankeropgaven maakten als onderdeel van de nationale optietoets (cel 2 in Tabel 1), Nederlandse leerlingen die de ankeropgaven maakten als onderdeel van de internationale toets (cel 5) en Vlaamse, Engelse en Deense leerlingen die de ankeropgaven maakten als onderdeel van de internationale wiskundetoets (cel 8).

Een beperking bij het beantwoorden van de eerste vraag is dat een cel 5 en cel 8 leerling slechts drie à zeven ankeropgaven heeft gemaakt. Deze leerlingen hebben ieder voor zich dus niet alle opgaven gemaakt en bovendien varieert het aantal leerlingen per ankeropgave. Dit is het gevolg van het toetsdesign dat bij TIMSS is gehanteerd en van de geselecteerde set ankeropgaven (zie Kuiper, Bos & Plomp, 1997). Dit gegeven levert een probleem op bij het nauwkeurig schatten van de ankerscore van deze groepen leerlingen, en dus bij de beantwoording van de eerste vraag.

Er doet zich ook nog een tweede, groter probleem voor. Verondersteld werd dat er geen verschillen in prestaties zouden optreden tussen de leerlingen in cel 2 en cel 5. Bij een eerste analyse is echter gebleken dat het percentage leerlingen dat de verschillende ankeropgaven correct heeft gemaakt voor de cel 2 leerlingen (nationale toets) in vijftien van de zestien gevallen *hoger* uitvalt dan die voor de cel 5 leerlingen (internationale toets). Bij vijf opgaven (onder andere bij V02 en R10) is het verschil zelfs 9% of meer. Slechts voor één opgave valt het percentage correct hoger uit voor de cel 5 leerlingen.

Hoe dit onverwachte resultaat te verklaren? De leerlingen die de nationale optietoets hebben gemaakt, komen uit dezelfde klassen als de leerlingen die de internationale toets hebben gemaakt. De systematiek op basis waarvan per klas toetsboekjes aan leerlingen zijn toegewezen, maakt het uiterst onwaarschijnlijk dat de cel 2 leerlingen op belangrijke achtergrondkenmerken (schooltypeniveau, leeftijd, sociaal-economische achtergrond) verschillen van de cel 5 leerlingen. De gevonden verschillen in percentages correct kunnen daar niet aan worden toegeschreven.

Ook verschillen in toetscondities zijn niet waarschijnlijk. De enige toegestane variatie in afnamecondities tussen de internationale wiskundetoets en de nationale optietoets betrof het gebruik van zakrekenmachine, geodriehoek, passer en dergelijke. Het gebruik ervan was niet toegestaan bij de internationale toets en bij het eerste deel van de nationale optietoets (met daarin de ankeropgaven en vier nationale opgaven), maar wel bij het tweede deel van de nationale optietoets (zestien nationale opgaven). Er is niets dat er op wijst dat genoemde hulpmiddelen bijvoorbeeld wel zijn gebruikt bij het maken van de (anker)opgaven in het eerste deel van de nationale optietoets. Bovendien zijn de ankeropgaven zodanig van aard dat het gebruik van genoemde hulpmiddelen leerlingen waarschijnlijk weinig of geen voordeel zou hebben geboden. Nagegaan is ook of het percentage correct voor een ankeropgave afhankelijk is van het internationale toetsboekje waarin het was opgenomen. Ook dat blijkt niet het geval. In welk internationaal toetsboekje een ankeropgave ook voorkwam, de percentages correct voor de opgave varieerden niet of slechts in geringe mate.

De meest waarschijnlijke oorzaak voor de gevonden verschillen is de *positie* van de verschillende opgaven in de beide toetsen. In de nationale optietoets kwamen de ankeropgaven voor in de volgorde zoals eerder beschreven. In de internationale toets zaten de ankeropgaven 'verpakt' tussen andere wiskunde- en science-opgaven. In de internationale toets werd een reeks van circa zes wiskunde-opgaven gevolgd door een reeks van circa zes science-opgaven. Het bij elkaar plaatsen van de zestien ankeropgaven heeft de cel 2 leerlingen, naar wordt aangenomen, in de gelegenheid gesteld deze opgaven meer geconcentreerd en dus beter te maken dan de cel 5 leerlingen. Gezien dit onvermoede effect ligt 't het meest voor de hand de cel 2 leerlingen bij een verdere vergelijking buiten beschouwing te laten en de beantwoording van de eerste onderzoeksvraag te beperken tot een vergelijking van de ankerscores van de cel 5 en de cel 8 leerlingen.

Hoe verhouden de prestaties van de cel 5 leerlingen op het anker zich nu tot die van de leerlingen in Vlaanderen,

Engeland en Denemarken? Is het inderdaad zo dat de verschillen in prestaties op het anker tussen de cel 5 en cel 8 leerlingen ongeveer in dezelfde orde van grootte liggen als de verschillen in prestaties op de internationale toets tussen de cel 6 en cel 9 leerlingen?

In tabel 2 staat onder meer weergegeven welke percentages leerlingen uit Nederland (cel 5), Vlaanderen, Engeland en Denemarken (alle cel 8) de ankeropgaven gemiddeld correct hebben gemaakt. De set ankeropgaven is het beste gemaakt door de leerlingen in Vlaanderen (elke ankeropgave correct gemaakt door gemiddeld 72,9% van de leerlingen die de opgave hebben gemaakt), gevolgd door de leerlingen in Nederland (69,3%), Engeland (63,1%) en Denemarken (59,7%). De gemiddelde scores van de vier groepen leerlingen op de internationale wiskundetoets in zijn geheel, gestandaardiseerd op een internationale schaal met een gemiddelde van 500 en een standaarddeviatie van 100, zijn: Vlaanderen 565, Nederland 541, Engeland 506 en Denemarken 502. De verschillen in gemiddelde wiskundescore tussen Vlaanderen en Nederland onderling en tussen Engeland en Denemarken onderling zijn niet significant. Leerlingen uit Vlaanderen en Nederland presteren wel significant beter op de internationale wiskundetoets dan leerlingen uit Engeland en Denemarken. Vergelijken we deze gemiddelde scores met de gemiddelde percentages correct scores op het anker, dan blijkt dat de beide groepen niet in volgorde van elkaar verschillen.

Relevanter is het echter om de scores op het anker te vergelijken met die op de *overige* opgaven uit de internationale wiskundetoets, dat willen zeggen de internationale wiskundetoets minus de zestien ankeropgaven. In tabel 2 staan voor de vier groepen leerlingen ook de scores op laatstgenoemde set van opgaven vermeld (eveneens in termen van gemiddelde percentages correct) in combinatie met de verschillen tussen deze scores en de gemiddelde percentages correct op het anker. Toepassing van een variantie-analyse wijst uit dat er grote verschillen bestaan tussen de scores van de leerlingen uit de vier landen op

Tabel 2: Gemiddelde percentages leerlingen in Nederland, Vlaanderen, Engeland en Denemarken die elk van de ankeropgaven respectievelijk de overige wiskunde-opgaven uit de internationale wiskundetoets correct hebben gemaakt

	m % correct Nederland n = 1954	m % correct Vlaanderen n = 2894	m % correct Engeland n = 1777	m % correct Denemarken n = 2344
ankeropgaven (cel 5 en 8)	69,3	72,9	63,1	59,7
overige internationale opgaven	65,5	72,7	54,9	53,9
verschil anker – overige	3,8	0,2	8,2	5,8

Noot: m % correct = gemiddeld percentage correct; n = aantal leerlingen

deze beide sets van opgaven. Voor de leerlingen in Denemarken, Engeland en Nederland zijn de scores op het anker significant hoger dan de scores op de set overige internationale opgaven, voor de leerlingen in Vlaanderen is er geen sprake van een significant verschil. Voor wat betreft de grootte van het verschil ligt Nederland (3,8%) duidelijk tussen enerzijds Vlaanderen (0,2%) en anderzijds Denemarken (5,8%) en Engeland (8,2%) in.

Dit algemene beeld wordt globaal ook per boekje gevonden. Dit betekent dat, hoewel de verschillen in scores op het anker en de overige internationale opgaven tussen de vier landen groot zijn, er geen grond is voor de veronderstelling dat de ankeropgaven voor leerlingen in Nederland relatief eenvoudiger zijn dan de overige internationale opgaven. Het verschil tussen de gemiddelde score op het anker en op de overige internationale opgaven ligt voor Nederland namelijk tussen die van de andere drie landen in.

Geconcludeerd kan worden dat de verschillen in prestaties op het anker tussen de cel 5 en cel 8 leerlingen in dezelfde lijn en in ongeveer dezelfde orde van grootte liggen als de verschillen in prestaties van deze vier groepen leerlingen op de overige opgaven uit de internationale wiskundetoets.

De tweede onderzoeksvraag

Bij de tweede onderzoeksvraag gaat het om een vergelijking van verschillen in prestaties van Nederlandse leerlingen in het tweede leerjaar voortgezet onderwijs op de nationale opgaven (cel 1) met die van de Nederlandse leerlingen in leerjaar 2 op de internationale wiskundetoets in zijn geheel (cel 6). Verondersteld wordt dat leerlingen beter zullen presteren op de nationale opgaven dan op de internationale wiskundetoets. Het was de bedoeling om de vergelijking tussen de cellen 1 en 6 uit te voeren via de cellen 2 en 5. Echter, het verschil in percentages correct op de zestien ankeropgaven tussen de cel 2 en cel 5 leerlingen en de onnauwkeurigheid van de ankerscore voor de cel 5 leerlingen (veroorzaakt door het feit dat niet

alle leerlingen alle ankeropgaven hebben gemaakt) hebben ook consequenties voor de mogelijkheid tot beantwoording van deze vraag.

Als eerste stap tot beantwoording van de tweede vraag is voor de leerlingen die de nationale optietoets hebben gemaakt, gekeken naar het verschil in gemiddelde percentage correct per leerling voor enerzijds de nationale opgaven (cel 1) en anderzijds de ankeropgaven (cel 2). Uit tabel 3 blijkt dat het gemiddelde percentage correct voor de set nationale opgaven, ongeacht schooltype en sekse, ongeveer 20% lager uitvalt dan voor het anker. Dat wil zeggen dat een leerling van de set nationale opgaven gemiddeld 20% minder correct maakt dan van de set ankeropgaven.

Dit verschil kan niet worden toegeschreven aan verschillen in inhoudelijke geschiktheid van de twee deelttoetsen. Uit analyse is namelijk gebleken dat de via beide deelttoetsen getoetste stof volgens verreweg de meeste leraren (86 à 89%) aansluit bij de stof die is behandeld in zowel VBO/MAVO-2 als HAVO/VWO-2. Een reden is mogelijk het verschil in *moelijkheidsgraad* van de twee deelttoetsen.

Uit verdere analyses blijkt dat een leerling die het anker goed maakt ook de nationale opgaven in het algemeen goed maakt (significante correlatie van .71; $p=.05$).

Betekent dat nu ook dat, in psychometrische termen geformuleerd, de set nationale opgaven en het anker één schaal vormen? Vormen beide ook samen met de internationale wiskundetoets één schaal?

Met behulp van schaalanalysetechnieken (Verhelst, Glas & Verstralen, 1995) is allereerst nagegaan of de set nationale opgaven, de set ankeropgaven en de set opgaven in de internationale wiskundetoets ieder afzonderlijk een schaal vormen (hetgeen het geval blijkt te zijn) en is vervolgens nagegaan of de set ankeropgaven één schaal vormt met enerzijds de set nationale opgaven (vergelijking cel 1 en cel 2) en anderzijds de opgaven in de internationale wiskundetoets (vergelijking cel 5 en cel 6). Uit laatstgenoemde analyse komt naar voren dat alle nationale opgaven, ankeropgaven en overige internationale op-

Tabel 3: Gemiddeld percentage correct beantwoorde nationale opgaven en ankeropgaven uit de nationale optietoets wiskunde, naar schooltypeleerjaar en geslacht

	m % correct vbo/mavo-2 n = 140		m % correct havo/vwo-2 n = 99	
	meisjes n = 60	jongens n = 80	meisjes n = 47	jongens n = 52
opgaven sets				
nationale opgaven in nationale optietoets wiskunde (cel 1)	40	45	64	67
ankeropgaven in nationale optietoets wiskunde (cel 2)	60	69	84	86

Noot: m % correct = gemiddeld percentage correct; n = aantal leerlingen

gaven tezamen één schaal vormen. Dit betekent dat een leerling die de nationale opgaven goed maakt ook goed gepresteerd zou hebben op de internationale wiskundetoets. Dit is de belangrijkste bevinding bij de beantwoording van de tweede vraag.

Conclusie

Is de door wiskunde-experts geuite twijfel over de representativiteit van de internationale wiskundetoets nu terecht geweest? Was er voldoende reden te vrezen voor onderschatting van de wiskundeprestaties van Nederlandse leerlingen op basis van deze toets?

De set nationale opgaven, het anker en de internationale wiskundetoets vormen, zoals we zojuist constateerden, één schaal. De nationale opgaven (cel 1) meten dus niet iets anders dan de internationale (cel 6). Dit alles demonstreert de toegevoegde waarde van de nationale opgaven. Op grond van het feit namelijk dat de drie sets opgaven één schaal vormen, kan logischerwijs worden geconcludeerd dat, indien de nationale opgaven ook gemaakt zouden zijn door leerlingen in andere landen (als onderdeel van de internationale toets), de verschillen in wiskundeprestaties tussen landen niet anders zouden zijn uitgevallen dan nu uit TIMSS is gebleken. Het feit dat de nationale opgaven (waarvan de inhoudelijke geschiktheid volgens experts en leraren boven elke twijfel verheven is) een goede schaal vormen met het anker en met de internationale wiskundetoets (waarvan de inhoudelijke geschiktheid door experts op voorhand sterk in twijfel werd getrokken, maar die – zo blijkt uit het nationale rapport – uiteindelijk toch heel redelijk is) toont aan dat *de internationale wiskundetoets goed bruikbaar is voor Nederland*. De set nationale opgaven lijkt weliswaar moeilijker dan de internationale wiskundetoets – het verschil in percentages correct voor de set nationale opgaven en het anker waarvan we eerder melding maakten, vormt hiertoe een indicatie – maar veel belangrijker is dat beide één schaal vormen. Daarmee lijkt er – gezien ook het feit dat de verschillen in prestaties op het anker tussen de cel 5 en cel 8 leerlingen in dezelfde lijn en in ongeveer dezelfde orde van grootte liggen als de verschillen in prestaties van deze vier groepen leerlingen op de overige opgaven uit de internationale wiskundetoets – geen reden te zijn om aan te nemen dat er sprake is van een onderschatting van wiskundeprestaties van Nederlandse leerlingen op basis van de internationale wiskundetoets. De ervaringen met de nationale optietoets maken evenwel duidelijk dat een replicatie van dit onderdeel (met aangepast anker) in combinatie met afname van de nationale optietoets in andere landen zeer relevant is.

Wilmad Kuiper, Klaas Bos en Tjeerd Plomp zijn verbonden aan de Universiteit Twente. Tjeerd Plomp is hoogle- raar binnen de vakgroep Curriculumtechnologie van de faculteit Toegepaste Onderwijskunde, Wilmad Kuiper universitair docent binnen diezelfde vakgroep en Klaas Bos senior onderzoeker bij OCTO. Tjeerd Plomp is sinds 1990 voorzitter van de IEA en trad op als voorzitter van de nationale begeleidingscommissie voor TIMSS populatie 2/3. Wilmad Kuiper is projectleider van TIMSS voortgezet onderwijs.

Literatuur

- Beaton, A.E., I.V.S. Mullis, M.O. Martin, E.J. Gonzales, D.L. Kelly, & T.A. Smith (1996). *Mathematics achievement in the middle school years. IEA's Third International Mathematics and Science Study*. Boston: Boston College.
- Commissie Ontwikkeling Wiskundeonderwijs (1992). *Trajectenboek Wiskunde 12-16*. Culemborg: Freudenthal Instituut/Educaboek.
- Eggen, T.J.H.M. (1993). Itemresponstheorie en onvolledige designs. In: T.J.H.M. Eggen & P.F. Sanders (Red.), *Psychometrie in de praktijk*, Arnhem: CITO, pp. 239-284.
- Kuiper, W.A.J.M., K.Tj. Bos & Tj. Plomp (1997). *Wiskunde en de natuurwetenschappelijke vakken in leerjaar 1 en 2 van het voortgezet onderwijs. Nederlands aandeel in TIMSS populatie 2*. Enschede: Universiteit Twente, OCTO.
- Lange, J. de (1995). Assessment: no change without problems. In: T.A. Romberg (Ed.), *Reform in school mathematics and authentic assessment*. Albany: State University of New York Press. pp. 87-172.
- Lange, J. & H. Boertien, (1994). *Model voor wiskundetoets nationale optie TIMSS 1995*. Enschede: Universiteit Twente, OCTO (interne notitie).
- Verhelst, N.D., C.A.W. Glas & H.H.F.M. Verstralen (1995). *One-parameter logistic model*. Arnhem: CITO.

Exemplaren van het rapport *Wiskunde en de natuurwetenschappelijke vakken in leerjaar 1 en 2 van het voortgezet onderwijs. Nederlands aandeel in TIMSS populatie 2* (à f 35,-) zijn te bestellen op het volgende adres: Universiteit Twente, Faculteitsbibliotheek TO Postbus 2177500, AE Enschede Tel.: (053) 4893588, Fax: (053) 4892895 E-mail: Nelissen@edte.utwente.nl