

Dit jaar werden de uitkomsten van de evaluatie van de basisvorming gepresenteerd. Naast de Inspectie, voerden ook het GION en het OCTO een onderzoek uit. In dit artikel gaat **Hans Kuyper** nader in op de resultaten van de wiskundetoets voor de cohorten '89 en '93.

Wiskunde in VOCL

Enkele maanden geleden is in de landelijke pers de nodige aandacht besteed aan de evaluatie van de Basisvorming. Begin september presenteerden onderzoekers van het Gronings instituut voor onderzoek van onderwijs, opvoeding en ontwikkeling (GION) en het Onderzoek Centrum Toegepaste Onderwijskunde (OCTO) hun resultaten op een door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) en het blad *Didaktief & School* georganiseerde persconferentie. Enkele weken later kwam de Inspectie met een integrale evaluatie. Vermoedelijk is in de beeldvorming vooral blijven hangen dat het met de Basisvorming niet zo goed is gesteld. De onderzoekers zelf oordelen daar heel wat genuanceerder over. In deze bijdrage ga ik op verzoek van de redactie in op een aantal resultaten die speciaal op wiskunde betrekking hebben.

Wat is VOCL?

VOCL staat voor 'Voortgezet Onderwijs Cohorten Leerlingen'. Een cohort is oorspronkelijk een eenheid in het Romeinse leger, bestaande uit even oude soldaten. In onder andere sociaal-wetenschappelijk onderzoek staat het veelal voor een 'jaarlichting'. In onderwijskundig onderzoek betreft het doorgaans (een steekproef van) de leerlingen die in een bepaald schooljaar in een bepaald leerjaar zitten. De VOCL-cohorten zijn mede opgezet ter evaluatie van de Basisvorming. Er is een pre-Basisvorming cohort bestaande uit circa 20.000 leerlingen die in schooljaar 1989/1990 in de eerste klas van het voortgezet onderwijs zaten (VOCL'89) en een Basisvorming cohort van eveneens ca. 20.000 leerlingen die in schooljaar 1993/1994 in de eerste klas zaten. Beide cohorten gelden als representatief voor de leerlingenpopulatie van de betreffende jaarlichting. In beide cohorten is in het derde leerjaar een wiskundetoets en een toets tekstbegrip Nederlands afgenomen bij de leerlingen die niet zijn blijven zitten.

De wiskundetoetsen

De wiskundetoets voor VOCL'89 is door het Cito ontwikkeld. Voor diverse details verwijst ik naar de bijdrage van

Boertien in Kremers (1990). Het volgende is daaruit een (korte) samenvatting. De toets diende geschikt te zijn voor de evaluatie van de Basisvorming en moest derhalve items bevatten behorend bij het in 1992 vigerende leerplan, items behorend bij het leerplan van de Basisvorming en items die tot beide leerplannen behoorden. In genoemde bijdrage wordt hierover opgemerkt:

'In het vigerende onderwijs zijn de opgaven die men aan de orde wil stellen, in het algemeen een goede afspiegeling van de opgaven waarmee men de leerstof geoefend heeft. Bovendien betreffen deze opgaven grotendeels formele wiskunde.' (p. 230)

En:

'Bij toepassingsgerichte wiskunde moet een leerling er blijk van geven in nieuwe situaties zelf zijn probleem te kunnen (her)definiëren, te vertalen en op te lossen.' (p. 231)

Dit heeft geleid tot vooral meerkeuze-vragen voor de toetsing van het (toen nog) vigerende onderwijs en overwegend open opgaven om te toetsen in hoeverre leerlingen de wiskunde van de (op dat moment toekomstige) Basisvorming beheersten. Een probleem was dat het tijdens de voorbereiding van de toetsconstructie (vanaf 1988) nog niet precies duidelijk was hoe het nieuwe leerplan eruit zou gaan zien. Tevens was het oorspronkelijk de bedoeling dat er één toets voor alle onderwijstypen (IHNO/ILTO tot en met VWO) zou komen – vermoedelijk vanuit de toen heersende 'ideologie van de Basisvorming'.

In totaal zijn 260 items/opgaven verzameld dan wel ontworpen. Iets meer dan de helft waren (voor het grootste deel bestaande) meerkeuze-vragen. De open opgaven zijn ontworpen door een team van drie docenten. Beide groepen opgaven hebben betrekking op vijf deelgebieden:

Re: voortgezet rekenen, algebra
Fu: functies, verbanden, grafieken
Me: meetkunde in R2 en R3
Ks: kansrekening en statistiek
Cx: complexe vaardigheden.

De items zijn verdeeld over tien toetsboekjes, die elk twaalf open en veertien meerkeuze-opgaven bevatten. Een aantal open opgaven bestond uit een a-vraag en een b-vraag. Het gemiddeld aantal opgave-delen per toets-

boekje was 35.5. De toetsboekjes zijn in een proefafname door in totaal bijna 8500 leerlingen – afkomstig uit het LHNO, LTO, MAVO, HAVO en VWO – gemaakt. De afname-tijd was twee lesuren, ofwel circa negentig minuten. De open opgaven zijn door een aantal geschoolde beoordelaars nagekeken, aan de hand van een beoordelingsfilosofie en een antwoordmodel met puntentoekening. Op grond van psychometrische analyses op de antwoorden op de meerkeuze-items en de puntentoeeningen op de open opgaven, aangevuld met beoordelingen van de opgaven door de docenten van de klassen die aan de proefafname deelnamen, is uiteindelijk één toets samengesteld. Deze toets bestaat uit zestien open vragen – met in totaal twintig onderdelen – en twaalf meerkeuzevragen; in totaal dus 32 opgave-onderdelen (hierna items genoemd). De genoemde onderdelen Re, Fu, Me, Ks en Cx zijn met respectievelijk acht, zeven, acht, zes en drie items vertegenwoordigd. Het vigerende leerplan is met zestien items vertegenwoordigd, het nieuwe met tien, terwijl de overige zes items op beide leerplannen betrekking hebben. De 32 items zijn volgens het scorevoorschrift goed voor een maximum van 140 punten. De ‘puntenwaarde’ verschilde tussen de items, waarbij het goed beantwoorden van moeilijke items meer punten opleverde dan het goed beantwoorden van makkelijke items.

Uit de resultaten van de proefafname bleek echter ook dat deze toets te moeilijk was voor de LHNO- en LTO-leerlingen. In een later stadium is daarom een eenvoudiger versie voor deze leerlingen (en die in het IHNO en ILTO) gemaakt. Tussen beide versies bestaat wel een overlap. De maximale score op de eveneens uit 32 opgaven bestaande IBO/LBO-versie bedraagt 80 punten. Beide toetsversies zijn in het voorjaar van 1992 afgenomen bij de leerlingen uit het 89-cohort toen die in de derde klas zaten.

Ten behoeve van de afname in het 93-cohort werd het wenselijk geacht de toetsen aan te passen. Hoe de besluitvorming hierover precies heeft plaatsgevonden, is niet geheel duidelijk. Vanuit een evaluatie-perspectief zou het optimaal zijn geweest als precies dezelfde twee toetsversies waren afgenomen, omdat dan immers tweemaal dezelfde meetlat zou zijn gebruikt. Vrijwel zeker om vakinhoudelijke redenen is tot de volgende aanpassing besloten. De informatie hierover is te vinden in het tweede deel van een interne Cito-publicatie (Boertien, Heuves & Kremers, 1995). Daaruit het volgende citaat:

‘Begonnen is met een globale inschatting te maken van welke opgaven van de twee toetsversies van VOCL ’89 het minst passen bij het nieuwe leerplan. Deze opgaven zijn geschrapt. Voor deze opgaven zijn zoveel mogelijk bij dezelfde domeinen waarop ze betrekking hadden vervangende opgaven geconstrueerd. De nieuwe opgaven zullen samen met de overgebleven opgaven in de VOCL ’89 toetsen een evenwichtige dekking van de leerstof geven.’

Kort samengevat bestond de VOCL ’89-toets dus uit opgaven die ten dele op het oude (toen vigerende) leerplan en ten dele op het nieuwe leerplan betrekking hadden, terwijl de VOCL ’93 toets vrijwel volledig uit items behorend

bij het nieuwe leerplan bestond. Deze onevenwichtigheid zou kunnen suggereren dat er vanuit het oude leerplan wel ‘transfer’ naar het nieuwe werd verwacht, terwijl het omgekeerde minder voor de hand liggend werd geacht.

Voor de constructie van de deels gewijzigde toetsversies is weer een proefafname georganiseerd, waarbij vier toetsboekjes aan bijna 1600 VBO-, MAVO- en VWO-leerlingen zijn voorgelegd. Op grond van de resultaten is een toetsversie voor het MAVO/HAVO/VWO gemaakt die uit 33 opgaven bestaat, die dit keer maximaal 101 punten konden opleveren, en een IVBO/VBO versie van eveneens 33 opgaven, goed voor maximaal 77 punten. In de nieuwe MAVO/HAVO/VWO versie zijn vijftien oude opgaven vervangen door zestien nieuwe. In de nieuwe IVBO/VBO versie zijn twee oude opgaven vervangen door drie nieuwe. Deze versies zijn in het voorjaar van 1996 afgenomen bij leerlingen uit het 93-cohort toen die in de derde klas zaten. Op een aantal scholen kon de afname echter pas na de zomervakantie plaatsvinden.

Equivalering van de toetsen

Er zijn dus in principe vier verschillende meetlatten gebruikt om het wiskundig niveau van de leerlingen in beide cohorten vast te stellen. Gelukkig biedt de toegepaste wiskunde een mogelijkheid om de resultaten, zowel die van beide cohorten als in elk cohort die van (1)VBO-leerlingen en MAVO/HAVO/VWO-leerlingen, toch vergelijkbaar te maken op een en dezelfde meetlat. Het vergelijkbaar maken (equivaleren) kan alleen indien er een aantal overlappende items (anker-items) is en geschiedt tegenwoordig met zogenaamde ‘item-response modellen’. In concreto is gebruik gemaakt van het ‘One Parameter Logistic Model’ (OPLM). De geïnteresseerde lezer verwijst ik naar een hoofdstuk van Engelen & Eggen in *Psychometrie in de praktijk* (Eggen & Sanders, 1993).

In het huidige geval zijn er zeven items die in alle vier toetsversies aanwezig zijn; in aanvulling daarop zijn er tien items die gemeenschappelijk zijn voor beide MAVO/HAVO/VWO versies en 23 items die gemeenschappelijk zijn voor beide (1)VBO versies. Daarnaast bestaat er nog een geringe extra overlap tussen de twee 89-versies (één item) en de twee 93-versies (twee items). Eén van de aan de equivaleringstechniek ten grondslag liggende assumpties is dat alle items één onderliggende vaardigheid meten. Opgaven die niet goed bij de andere opgaven blijken te passen, kunnen beter worden weggelaten. In het ideale geval is de rangordening van de moeilijkheid van de anker-items in de verschillende toetsversies hetzelfde. Anker-items die relatief veel inversies opleveren, kunnen eveneens beter worden verwijderd. Door in een min of meer iteratief proces bepaalde wegingsfactoren aan te passen en/of instabiele items weg te laten, wordt toegewerkt naar een optimaal resultaat. In het huidige geval bleek het raadzaam drie items weg te laten. Hiervan behoorde er één tot het gemeenschappelijke anker en twee tot het anker van de MAVO/HAVO/VWO-versies. Het resul-

taat van de equivaleringsprocedure bestaat onder andere uit een schatting van de vaardigheid van elke leerling die één van de vier toetsversies heeft gemaakt. Een ander resultaat is een schatting van de betrouwbaarheid (Cronbach's α) voor elk van de vier toetsversies. De betrouwbaarheidswaarden zijn respectievelijk .86 en .84 voor de MAVO/HAVO/VWO-versie van het 89-cohort en het 93-cohort en .87 voor beide (I)VBO-versies. Dit zijn hoge waarden, zeker indien in aanmerking wordt genomen dat het aantal items niet bijzonder hoog is en diverse onderdelen binnen de wiskunde worden gedekt. De vaardigheidsschattingen van de leerlingen zijn getransformeerd naar T-scores, die een gemiddelde van 50.0 hebben en een standaard deviatie van 10.0.

Resultaten

De begin september in *Didaktief & School* gepubliceerde resultaten voor wiskunde zijn in onderstaande tabel samengevat.

	89-cohort	93-cohort	verschil
IVBO	32.8	33.8	+1.0
VBO	43.5	44.3	+0.8
MAVO	50.4	51.2	+0.8
HAVO	55.9	55.6	-0.3
HAVO/VWO	58.7	58.9	+0.2
VWO	61.8	60.2	-1.6
overall	49.7	50.4	+0.7

Tabel 1 Cohortvergelijking T-scores wiskunde per onderwijstype

Andere dan de in de tabel opgenomen klastypen waren in minstens één van beide cohorten onvoldoende vertegenwoordigd. Dat het gemiddelde van de twee overall cohort-gemiddelden niet exact 50.0 is, vindt zijn oorzaak in de omstandigheid dat in het 89-cohort meer leerlingen aan de toetsafname hebben deelgenomen dan in het 93-cohort (10.800 vs. 8100). Overall is er een geringe toename van 0.7 te zien op de geëquivalente wiskundetoets – of van de gemeten wiskundevaardigheid. Bij nadere beschouwing is in de drie laagste onderwijstypen enige vooruitgang te zien, terwijl in het hoogste onderwijstype zich een wat grotere afname manifesteert. De leerlingen in HAVO- en HAVO/VWO-klassen zijn vrijwel gelijk gebleven. Op deze manier bekeken is er weinig aanleiding tot ongerustheid over wiskunde in de Basisvorming en baart alleen het VWO enige zorg. Voordat verdere bespiegelingen zinvol zijn, moet eerst worden ingegaan op de vergelijkbaarheid van de twee cohorten.

Vergelijkbaarheid van de cohorten

Aan de vergelijkbaarheid zijn drie aspecten te onderscheiden. In de eerste plaats is het de vraag of (of: in hoeverre) de twee cohorten bij aanvang gelijkwaardig waren. Om deze vraag te beantwoorden, is een vergelijking gemaakt op enkele achtergrondkenmerken van de leerlingen, namelijk het 'advies eind basisonderwijs', sekse, sociaal-economische status (SES) en etnische achtergrond en op enkele in het eerste leerjaar afgenomen toetsen/tests, namelijk drie 'entreetoetsen' en twee subtests voor non-verbale intelligentie. De volgende informatie is ontleend aan tabellen 2.1 en 4.1 van het rapport *Onderwijsresultaten van VOCL'89 en VOCL'93 leerlingen* (Van der Werf, Lubbers & Kuiper, 1999).

Het gemiddelde advies is in beide cohorten precies gelijk. In beide cohorten zijn de jongens in enige mate in de meerderheid, maar in het 93-cohort een half procent minder. Als indicator voor SES is in onderwijsonderzoek de hoogste door de ouders/verzorgers gevolgde opleiding het meest gebruikelijk en het meest geschikt. Op dit aspect is een toename te zien van 0.3 op een schaalbereik van 5.0. De ouders van de leerlingen in het 93-cohort zijn gemiddeld iets hoger opgeleid. Het percentage allochtone leerlingen is met bijna 1% toegenomen; ruim de helft hiervan komt op rekening van Turkse leerlingen. Op de entreetoetsen (Nederlandse taal, rekenen, informatieverwerking) is een geringe toename en op de intelligentie subtests is een vrij grote toename te zien. De hierboven genoemde verschillen zijn significant ($p < .001$), maar vanwege de zeer grote aantallen zegt dat niet zoveel. De toenames in SES en op non-verbale intelligentie zijn met effectgroottes van .30 à .40 relevant.¹

In de tweede plaats is, vrijwel zeker onder invloed van de Basisvorming, het zittenblijven in de onderbouw vrij aanzienlijk afgenomen (van ongeveer 12% tot ongeveer 6%). Tevens zijn zowel de opstroom naar een hoger onderwijstype dan geadviseerd als de afstroom naar een lager onderwijstype toegenomen. De toename van de afstroom is daarbij groter dan de toename van de opstroom.

In de derde plaats zijn in het derde leerjaar aanzienlijk minder leerlingen getoetst dan de circa 20.000 cohort-leerlingen. De voornaamste reden hiervoor is dat veel scholen niet wilden meewerken aan de betreffende toetsafname, onder andere vanwege het late tijdstip in het schooljaar. Een tweede reden is het zittenblijven. Ten slotte waren er leerlingen naar een andere school gegaan, waren er al voortijdig schoolverlaters en waren er leerlingen bij de toetsafname absent. Vooral in het 93-cohort is de uitval van scholen groot. De combinatie van deze factoren heeft tot de reeds eerder genoemde aantallen (10.800 en 8100) geleid. Dit roept de vraag op in welke mate deze in leerjaar 3 getoetste leerlingen van beide cohorten vergelijkbaar zijn. De informatie hierover staat in tabel 2.3 van het eerder genoemde rapport (Van der Werf, Lubbers & Kuiper, 1999). Het blijkt dat de in het derde leerjaar van het 93-cohort getoetste leerlingen gemiddeld

een hoger advies hebben en gemiddeld hoger scoorden op de in het eerste leerjaar afgenomen entree-toetsen en intelligentie subtests dan de in het derde leerjaar van het 89-cohort getoetste leerlingen. Op intelligentie is het verschil relevant: een effectgrootte van ongeveer .30.

Verdere uitwerking

In sommige opvattingen moet bij de interpretatie van de in tabel 1 gepresenteerde gegevens nog gecorrigeerd worden voor de achtergrondkenmerken van de leerlingen, met name voor kenmerken die samenhangen met de score op de wiskundetoets. De drie hierboven genoemde variabelen (advies, entree-toets en non-verbale intelligentie) vertonen substantiële correlaties met de wiskundescore: in het 89-cohort en het 93-cohort bedraagt de correlatie met het advies respectievelijk .73 en .70, met de score op de entree-toets .76 en .76 en met de intelligentiescore .31 en .41. Het geringe overall verschil ten gunste van het 93-cohort kan derhalve veroorzaakt zijn doordat de betreffende leerlingen gemiddeld hogere scores hebben op deze cognitieve variabelen.

Deze veronderstelling is in een reeks meerniveau-analyses uitgewerkt. Per saldo resulteert dan een verwaarloosbaar (niet eens significant) verschil ten nadele van het 93-cohort. Zie hiervoor tabel 4.20 in het eerder genoemde rapport (Van der Werf, Lubbers & Kuyper, 1999).

Ten slotte wil ik ingaan op een in mijn ogen opvallend patroon, dat ik pas bij de voorbereiding van deze bijdrage heb ontdekt en dat dus nog niet eerder is gepubliceerd. Het gaat hierbij om gedetailleerde analyses op item-niveau, en wel van de p-waarden, voordat in de equivalentieprocedure herparametriseringen hebben plaatsgevonden. [Wellicht geheel overbodig: de p-waarde van een meerkeuze-item is de proportie leerlingen die het juiste antwoordalternatief heeft omcirkeld en van een open opgave het gemiddelde toegekende aantal punten gedeeld door het maximum aantal punten dat voor dat item mogelijk is. De p-waarde wordt doorgaans de 'item moeilijkheid' genoemd, hoewel 'item makkelijkheid' correct zou zijn.]

Ik beperk me voornamelijk tot de MAVO/HAVO/VWO toetsversies. Nog een opmerking vooraf: in de versie van het 89-cohort is de score op twee deelopgaven samengenomen, aangezien uit controle van de datafile bleek dat sommige beoordelaars de totale puntentoekenning in één veld van het instreepformulier hadden geplaatst.

Zoals eerder opgemerkt, bestond de 89-toets ten dele uit items van het destijds vigerende leerplan, ten dele uit items van het nieuwe leerplan en ten dele uit gemeenschappelijke items en zijn uit de 93-toets de items van het oude leerplan weggelaten. Het is daarom interessant om bij de leerlingen van het 89-cohort te kijken naar de gemiddelde p-waarde van de overlappende en de verwijderde items en evenzo om bij de leerlingen van het 93-cohort te kijken naar de gemiddelde p-waarde van de overlappende en de nieuw toegevoegde items. Deze vergelijking

wordt gemaakt in tabel 2.

	89-cohort (oude leerplan)	93-cohort (nieuwe leerplan)
overlappende items (17)	.565	.617
verwijderde items (14)	.440	–
toegevoegde items (16)	–	.580

Tabel 2 Gemiddelde p-waarde van drie groepen items in twee cohorten

Hier doet zich het opmerkelijke verschijnsel voor dat de veertien (of vijftien – zie de opmerking over het samenemen van twee deelopgaven) verwijderde items, die overwegend betrekking hadden op het voor de leerlingen van het 89-cohort vigerende leerplan (formele wiskunde) door deze leerlingen gemiddeld behoorlijk wat (.125) slechter zijn gemaakt dan de zeventien items die niet zijn verwijderd, omdat ze al (tevens) tot het nieuwe leerplan behoorden. Ook blijkt dat deze zeventien [(tevens) tot het nieuwe leerplan behorende] items door de leerlingen van het 93-cohort, die met dat leerplan zijn onderwezen, .052 beter zijn beantwoord dan door de leerlingen die met het oude leerplan zijn onderwezen. Ten slotte blijkt dat de zestien nieuw toegevoegde items door de leerlingen van het 93-cohort .037 slechter zijn gemaakt dan de zeventien overlappende items. Het resultaat is dat de 93-toets in termen van de gemiddelde p-waarden makkelijker is dan de 89-toets (.599 vs. .508). Dit verschil wordt door de equivalentieprocedure dus vrijwel volledig rechtgetrokken. De (1)VBO-versies hebben dertig overlappende items. De gemiddelde p-waarde op deze items is .559 in het 89-cohort tegen .581 in het 93-cohort – een bescheiden toename van .022. De twee weggelaten items hadden in het 89-cohort p-waarden van .542 en .363. De drie toegevoegde items hebben in het 93-cohort p-waarden van .542, .191 en .109. Deze laatste twee items zijn verreweg de moeilijkste van de gehele toets.

Tot zover de feitelijke resultaten. De vraag is wat hetgeen ik als 'opmerkelijk verschijnsel' heb gekarakteriseerd, zou kunnen betekenen. Het is mijns inziens vooral aan de vakdidactici om deze vraag te beantwoorden. Niettemin waag ik me aan een interpretatie van mijn kant. Onder de assumptie dat de items een goede afspiegeling vormen van de leerdoelen die met beide leerplannen beoogd werden/worden, of – meer precies – dat de items die op het oude leerplan betrekking hadden niet wat moeilijkheid betreft gebiased waren ten opzichte van de items van het nieuwe leerplan, lijkt het er sterk op dat het oude leerplan moeilijker was dan het nieuwe. Ofwel: de meer formele wiskunde was voor de gemiddelde leerling in de onder-

bouw van het MAVO/HAVO/VWO moeilijker dan de 'nieuwe' contextrijke wiskunde, maar leidde wel tot 'transfer' bij het oplossen van opgaven behorend tot het nieuwe leerplan.

Hierbij nog enkele kanttekeningen. In de eerste plaats gaat deze interpretatie verder dan op grond van de resultaten op de geëquivalenteerde toets verantwoord is. In de tweede plaats betreft het alleen de vergelijking met de eerste lichting van leerlingen in de Basisvorming. In de derde plaats is het achteraf niet geheel eenduidig welke items van de 89-toetsversie tot welk leerplan behoren. En in de vierde plaats is het erg jammer dat vrijwel alle tot het oude leerplan behorende items zijn weggelaten uit de 93-toetsversie. Daarin is vrijwel zeker maar één 'oud' item opgenomen, namelijk een meerkeuze-item:

$$p(p-5) = \dots [2p-5p; p^2-5p; 2p-5; p^2-5]$$

In het 89-cohort is de p-waarde .783, tegen .635 in het 93-cohort.

Er is nog één ander overlappend item dat door de leerlingen van het 89-cohort beter is gemaakt dan door de leerlingen van het 93-cohort. Dat luidt als volgt:

Een lift is 1 meter breed, 1 meter diep en 2 meter hoog.
Kan er een stevige ijzeren pijp in die 2.5 meter lang is?
Licht je antwoord toe.

In het 89-cohort is de p-waarde .230, tegen .195 in het 93-cohort.

Er zijn vier overlappende items die door de leerlingen in het 93-cohort aanzienlijk beter zijn gemaakt dan door de leerlingen in het 89-cohort:

- item 19 (tegels) uit de 93-versie, met p-waarden van .278 en .503
- item 25 (raket) uit de 93-versie, met p-waarden van .500 en .625
Dit item is ook vertegenwoordigd in de twee (1)VBO-versies; de p-waarden daar zijn respectievelijk .209 en .355
- item 14 (staatsbosbeheer) uit de 93-versie, met p-waarden van .387 en .504
- item 13 (waarde van een auto) uit de 93-versie, met p-waarden van .600 en .702.

Ten slotte nog een verdergaande speculatie. Eerder dit jaar is in de landelijke pers aandacht besteed aan de afname in de keuze van exacte vakken, onder andere ook bij leerlingen van deze twee VOCL-cohorten. Bij de leerlingen die onvertraagd waren doorgestroomd naar HAVO-5 en VWO-5 bleek de keuze van wiskunde B als examenvak met respectievelijk 7% en 8% te zijn afgenomen. Hier tegenover staat een toename in de keuze van wiskunde A

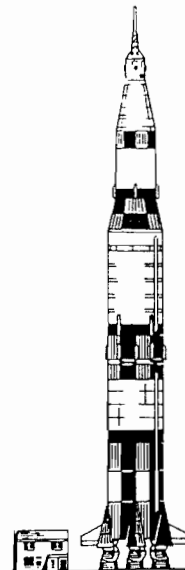
RAKET

Op het plaatje hiernaast zijn een huis en een raket van opzij gezien op dezelfde schaal getekend.

Hoe hoog schat je het huis in werkelijkheid?

Benader door meten en redeneren de werkelijke lengte van de raket.

Licht je antwoord toe.



Item 25 (raket) uit de VOCL'93 Toets wiskunde Versie A

met respectievelijk 8% en 11%. Destijds heb ik me als mogelijke verklaring laten ontvallen dat de exacte vakken misschien niet erg 'sexy' zijn. Een andere mogelijkheid die zich op grond van de huidige gegevens opwerpt, is dat de aansluiting van wiskunde A op de wiskunde van de Basisvorming beter is dan de aansluiting van wiskunde B.

Hans Kuyper, GION, Gronings instituut voor onderzoek van onderwijs, opvoeding en ontwikkeling, Rijksuniversiteit Groningen

H.Kuyper@ppsw.rug.nl

tel. 050-3636653

Noot

[1] De effectgrootte is gedefinieerd als het verschil tussen twee gemiddelden, gedeeld door de 'gepoolde' standaard deviatie.

Literatuur

Boertien, H., T. Heuves & E.J.J. Kremers (1995). *Constructie toetsen Nederlands en wiskunde voor VOCL-cohortonderzoek 1993*. Arnhem: Cito.

Eggen, T.J.H.M. & P.F. Sanders (1993). *Psychometrie in de praktijk*. Arnhem: Cito.

Kremers, E.J.J. (1990). *Overzicht van leerresultaten aan het einde van de eerste fase voortgezet onderwijs*. Arnhem: Cito.

Werf, M.P.C. van der, M.J. Lubbers & H. Kuyper (1999). *Onderwijsresultaten van VOCL'89 en VOCL'93 leerlingen*. Groningen: GION.