

In eerdere nummers van de Nieuwe Wiskrant werd door Cito-medewerkers een onderzoek beschreven naar de vergelijkbaarheid van open en gesloten (meerkeuze) vormen van toetsvragen. Ze komen tot de conclusie 'dat open en meerkeuzevragen uitwisselbaar zijn in de zin dat zij een beroep doen op dezelfde wiskundige vaardigheden'. **Henk van der Kooij** plaatst kritische kanttekeningen bij deze conclusie en bij andere bevindingen van dit onderzoek.

Wie bepaalt wat goed is voor mijn leerling?

Docenten wiskunde denken genuanceerder dan (sommige) Cito-medewerkers

Inleiding

Sinds de privatisering is Cito niet meer het door de overheid gesubsidieerde instituut dat onder andere de productie van eindexamens verzorgt. Als commerciële instelling probeer je natuurlijk werk te krijgen op het gebied waar je deskundig bent. Daarbij maak je gebruik van de expertise die in huis aanwezig is. Dat is prijzenswaardig. Veel minder begrijpelijk is het gebruiken van vermeende expertise om het onderwijsveld van diensten te voorzien waaraan mijns inziens geen behoefte is. Een gevoel van totaal onbegrip bekruipt mij bij het lezen van de twee artikelen over het Volg en AdviesSysteem (VAS) van twee Cito-medewerkers (Kuhlemeier & Steentjes, 2002a, 2002b). In beide artikelen worden 'open' en 'gesloten' vraagstellingen met elkaar vergeleken. Wiskundedocenten is naar hun voorkeur gevraagd (eerste artikel), open en gesloten versies van eenzelfde toets zijn aan verschillende groepen leerlingen voorgelegd (artikel 2) en de resultaten van beide acties worden geanalyseerd. De belangrijkste uitkomst van het onderzoek is de conclusie die ook al in de kop staat vermeld: open en meerkeuzevragen zijn uitwisselbaar in de zin dat zij een beroep doen op dezelfde wiskundige vaardigheden. En dat is precies de conclusie die voor Cito rechtvaardigt dat ze kan doorgaan met het afnemen van toetsen in multiple-choice format. Dat blijkt dan ook de enige doelstelling van het onderzoek: de rechtvaardiging van een al gemaakte keuze. Maar grote delen van de twee artikelen zijn strijdig met deze conclusie. Dat zal ik in het vervolg proberen duidelijk te maken, nadat eerst iets wordt gezegd over de zin van toetsen, het begrip wiskundige vaardigheid en het verschil tussen open en gesloten vraagstellingen.

De zin van toetsen

In het eerste artikel onderscheiden de auteurs drie doelen van toetsing: voortgangscontrole, resultaatbepaling en determinatie. Bij de eerste categorie wordt gezien aan leerlingen staan in het leerproces met het doel tekortkomingen in het leren (en volgens mij zou daar ook heel goed bij passen: en/of in het onderwijzen!) zichtbaar te

maken. De tweede is de, in de onderwijspraktijk nog steeds meest gebruikte, vorm van hoofdstuktoetsing en schriftelijk examen. Dat deze strikte scheiding tussen de twee eerste vormen nog steeds door Cito wordt gehanteerd is wellicht verklaarbaar vanuit internationaal toetsdeskundigheidsperspectief, maar in Nederland wordt daarover sinds de invoering van wiskunde A in het VWO wat genuanceerder gedacht. Het nieuwe examenprogramma van toen (1985) bracht de discussie over doel en vorm van toetsing op gang. In zijn proefschrift (De Lange, 1987) hanteert Jan de Lange vijf essentiële criteria voor toetsing, waaronder de toetsvormen die hij toen nog 'alternatief' noemde:

- toetsen moeten bijdragen aan het leerproces
- leerlingen moeten de kans krijgen te laten zien wat ze kennen en kunnen (positief toetsen) en niet wat ze niet kennen of kunnen (negatief toetsen)
- de doelen van het vak moeten in de toetsing zoveel mogelijk geoperationaliseerd worden
- de kwaliteit van een toets wordt niet in de eerste plaats bepaald door de mogelijkheid tot objectieve scoring
- de toets moet binnen de schoolpraktijk passen.

De termen 'voortgangscontrole' en 'resultaatmeting' passen niet zo goed bij de geest van deze criteria: ze zijn te zeer gericht op het meten van eventuele tekortkomingen. Met name de eerste drie van de genoemde criteria geven aan dat toetsing een geïntegreerd onderdeel vormt van het totale leerproces. De toetsing geeft van en voor iedere individuele leerling een beeld van zijn/haar groei in wiskundige kennis en vaardigheid gedurende de hele schoolloopbaan. En dat is dan ook precies het doel van toetsing: een eerlijk beeld geven van de vorderingen die een lerende maakt. Daarbij gaat het in Nederland al geruime tijd niet meer alleen om procedurele kennis ofwel *product*vaardigheden (feiten, algoritmen, reproductie), maar ook om *proces*vaardigheden (redeneren, generaliseren, reflecteren, transfer). Het begrip wiskundige vaardigheid wordt daarom ook op verschillende niveaus gedefinieerd; daarover straks meer. Omdat procesvaardigheden moeilijk toetsbaar zijn op de manier van de gewone proefwerken en schriftelijke examens, worden tegenwoordig ook andere toetsvormen in het Nederlandse

onderwijs gebruikt. Aan profielwerkstukken, computerpractica en praktische opdrachten wordt individueel of in kleine groepjes gewerkt. Ook de Wiskunde A-lympiade en de Wiskunde B-dag (voor HAVO- en VWO bovenbouw) worden op veel scholen voor dit doel ingezet. Met het gebruiken van verschillende vormen van toetsing wordt een breed beeld geschetst van de vaardigheden die een leerling tijdens het leerproces ontwikkelt. En juist zo'n breed beeld zou doorslaggevend moeten zijn bij de advisering over mogelijke vervolgleertrajecten van elke individuele leerling. Daarmee zijn we aangekomen bij wat de schrijvers van de twee artikelen het derde doel van toetsing noemen: de determinatie. 'De wiskundetoetsen in het VAS systeem van Cito laten zien voor welke leerweg (in het VMBO) een leerling het meest geschikt is.' (letterlijk citaat uit Kuhlemeier & Steentjes, 2002a).

In het licht van het bovenstaande en het vervolg lijkt het op zijn minst verdacht om te pretenderen dat te kunnen doen door leerlingen een setje van achttien meerkeuzevragen voor te leggen waarin alleen de wiskundige *product*vaardigheden worden getoetst.

Wiskundige vaardigheden

De term procesvaardigheden deed zijn intrede in het Nederlandse wiskundeonderwijs bij de invoering van wiskunde A in het VWO (HEWET-project). Voor het operationaliseren in de toetsing van *alle* doelen van het nieuwe programma (zie de criteria hierboven) kwam De Lange tot het definiëren van wiskundige vaardigheden op verschillende niveaus, die als zodanig ook in toetsvragen aan bod moeten komen (De Lange, 1987). In de loop van de tijd is de driedeling verder aangescherpt. In de nationale optie van de TIMSS-toetsen werden door De Lange drie niveaus van wiskundige vaardigheden gebruikt om de verschillende soorten toetsvragen te onderscheiden. Daarbij werd de toetspiramide ontworpen als visueel middel om de verschillende dimensies van toetsvragen mee te typeren. In het internationale PISA-project zijn deze drie niveaus van wiskundige vaardigheden overgenomen (OECD, 2003):

1. Reproductie, definities en (directe) berekeningen.
2. Verbanden en integratie, nodig voor probleem solving.
3. Mathematiseren, wiskundig denken, generalisatie en inzicht.

Het eerste niveau bevat alleen de productvaardigheden. Bij de niveaus 2 en 3 zijn, naast beheersing van de productvaardigheden, hogere orde vaardigheden in het geding. Daar moet je, om een probleem op te lossen, eerst aan de gang met denken over de wiskundige aanpak van het probleem, (dwars)verbanden zien of kunnen switchen tussen verschillende representatievormen. Op niveau 2 wordt de leerling geacht zijn eigen wiskundegereedschappen te kiezen. Op niveau 3 word je ook geacht wiskundige kennis te gebruiken binnen een geheel nieuwe context, te generaliseren en te bewijzen. Veel uitgebreider staan deze vaardigheden met voorbeelden beschreven

in twee Engelstalige publikaties (De Lange, 1999; Dekker & Querelle, 2001). Toetsvragen op niveau 1 kunnen (eventueel) in meerkeuzevorm worden gesteld; vragen op niveau 2 en 3 zijn absoluut niet in te passen in het keurslijf van de meerkeuze, omdat zulke vragen altijd beargumenteerde beantwoording vereisen. In feite vervalt voor mij daarmee al direct het belang van het onderzoek van Kuhlemeier en Steentjes. Het heeft geen zin om een vervolgtraject voor leerlingen vast te stellen op alleen maar toetsvragen van niveau 1. Als de wiskundetoets binnen VAS al iets toetst, dan gaat het alleen maar om de productvaardigheden. En hoewel beheersing van die vaardigheden natuurlijk nodig is, heeft dat op zich niet zoveel voorspellende waarde over het met succes vervolgen van welk leertraject dan ook. Daarmee zou ik de reactie op de twee artikelen kunnen besluiten, maar er zijn meer zaken die mij niet bevallen. De (vermeende) tegenstelling tussen 'gesloten' en 'open' is een nadere beschouwing waard, maar ook de manier waarop gegevens uit de onderzoeken onder docenten en leerlingen worden geanalyseerd roept om commentaar.

Open of gesloten?

In het circuit van toetsdeskundigen wordt de term 'open vraag' nog steeds gedefinieerd als ontkenning van een 'gesloten vraag'. Omdat een gesloten vraag in dat circuit equivalent is met een meerkeuzevraag, worden dus alle vragen waarbij niet een set antwoorden is meegeleverd waaruit een leerling het juiste alternatief moet kiezen, bestempeld als 'open'. Dat kan hoogst verwarrend zijn, omdat in wiskundeonderwijskringen de termen open en gesloten meestal op een heel andere manier worden gebruikt. Bij een gesloten probleemstelling ligt de te volgen oplossingsstrategie al besloten in de vraagstelling; bij een open probleemstelling wordt van de leerling verwacht dat hij zijn oplossingsstrategie zelf bepaalt. Gesloten vragen zijn in deze zienswijze dus bijna per definitie vragen op niveau 1 (feiten, definities en directe berekeningen). Een vraag als 'bereken $36^2 + 26^2$ ', die door Kuhlemeier en Steentjes wordt bestempeld als open (omdat er geen setje antwoorden bijstaat waaruit moet worden gekozen) is een volledig dichtgetimmerd direct commando dat moet worden uitgevoerd: meer gesloten is schier onmogelijk!

Ook de andere acht voorbeelden van toetsvragen die worden genoemd hebben deze vermeende tegenstelling tussen open en gesloten. Het is dan ook niet verbazingwekkend dat het merendeel van de ondervraagde docenten vindt dat bij een aantal van de opgaven in beide versies dezelfde wiskundige vaardigheid wordt gemeten. Er worden immers directe berekeningen gevraagd en er is geen verschil tussen het geven van een enkel getalsmatig antwoord zonder een rekenstap (zoals bij bovenstaande opgave) en het aankruisen van de juiste letter. Maar er zijn ook andere vragen in het lijstje van negen opgaven waarbij het wel degelijk verschil maakt. En daarover denken

docenten inderdaad veel genuanceerder dan deze Cito-medewerkers (en niet, zoals zij suggereren met de kop van hun eerste artikel, genuanceerder dan de lezer van de Nieuwe Wiskrant)! Zo vinden de docenten terecht dat een ‘open’ vraag waar een exact antwoord moet worden berekend een beroep doet op andere wiskundige vaardigheden, dan het in de ‘gesloten’ vorm kiezen in welk van de vier gegeven intervallen het juiste antwoord zich bevindt. Ook zijn ze (gelukkig!) van mening dat een opgave waarin rekenstappen moeten worden gemaakt om bij het antwoord te komen, in de open vorm meer informatie geeft over de kennis en kunde van een leerling dan wanneer alleen het goede eindantwoord moet worden aangekruist.

Verstandige docententaal!

‘Gemiddeld eenderde van de ondervraagde docenten is van mening dat de open versie een beroep doet op andere wiskundige vaardigheden dan de meerkeuzeversie’ staat in het eerste artikel. Waarom statistische beschouwingen over de uitkomsten van de docentenvragenlijst niet relevant zijn, wordt verderop beschreven. Het gaat me nu even om de toelichting die van deze docenten werd gevraagd. Ik lees onder andere de volgende voorbeelden van verstandige docententaal:

- bij de open versie wordt een leerling gedwongen zijn overwegingen logisch te verwoorden; bij de meerkeuzeversie weet ik niet of het een schatting, een gok of een berekening is
- de gesloten versie geeft geen enkel inzicht in wat de leerlingen eventueel fout doen.

Dit lijken mij heel gezonde reacties die ook goed passen bij het doel van de VAS: het beargumenteerd aangeven van een mogelijk vervolgleertraject voor een individuele leerling. Het stuit mij dan ook heftig tegen de borst als je de mening van de Cito-medewerkers leest:

‘Verbazingwekkend is wellicht dat veel docenten nog steeds een voorkeur hebben voor open vragen als het gaat om belangrijke beslissingen. Voor het type plaatsingsbeslissingen, waarvoor het Volg- en AdviesSysteem van het Cito in de markt is gezet, geeft 42% de voorkeur aan open vragen. Voor deze docenten weegt het voordeel van rijkere informatie over het oplossingsproces *kennelijk nog steeds* op tegen de extra werklast die het nakijken van open vragen met zich meebrengt.’

Eigenlijk worden docenten die het goede willen voor hun leerlingen hier neergezet als domme tijdverspillers, die geen oog hebben voor efficiëntie en time-management! We kunnen ons gelukkig prijzen met onderwijsexperts in het veld die nog de tijd nemen om hun leerlingen goed beargumenteerd en op eigen gezag over vervolgleertrajecten te adviseren, ondanks alle administratieve en organisatorische bijkomstigheden waarmee ze in het hedendaagse onderwijs worden opgezadeld. De tijd die ze winnen door niet deel te nemen aan zo’n administratieve VAS-constructie kan dan veel beter wiskundig inhoudelijk worden benut.

Hoe praat je recht wat krom is?

Het Cito-onderzoek is opgezet om een reeds genomen beslissing te rechtvaardigen. Die beslissing om een toets met alleen open vragen te vervangen door één met alleen meerkeuzevragen, was gebaseerd op een klacht van schoolleiders dat het nakijken van open vragen te veel tijd in beslag nam. Het zou verstandiger zijn geweest om een inhoudelijk onderzoek naar de vergelijkbaarheid te doen voordat blind werd toegegeven aan de administratieve en time-management druk. Nu restte kennelijk niet meer dan een poging om achteraf aan te tonen dat die beslissing niet hoeft te worden heroverwogen. En dat is niet zo moeilijk: het is een kwestie van met getallen strooien en daarbij alleen die getallen gebruiken die in de kraam te pas komen. Een paar voorbeelden wil ik u niet onthouden. In de tabel worden per opgave twee getallen getoond: het percentage docenten dat meent dat dezelfde vaardigheden worden getoetst en het percentage docenten dat voorkeur uitspreekt voor de open versie. Dat is geen volledige informatie, want de niet-getoonde percentages bestrijken nog twee categorieën: de andere mening of *geen* mening. Zo lijkt het erop dat, als 37% voorkeur uitspreekt voor de open versie van opgave 5, de andere 63% voorkeur heeft voor de gesloten versie. In werkelijkheid is het aantal voorstanders van de gesloten versie bij deze opgave lager dan 37%. Dat kan ik afleiden uit een enkel zinnetje dat wordt gewijd aan opgave 7. Dat is de enige opgave waar de gesloten versie het wint van de open versie (40% tegen 38% en dus kennelijk 22% ‘geen mening’!). Als de voorkeuren voor de meerkeuzeversie waren vermeld in plaats van de voorkeuren voor de open versie, dan was er dus bij geen enkele vraag een meerderheid van docenten gevonden met voorkeur voor de gesloten versie. Zo kunnen op zichzelf kloppende getallen toch een heel verkeerd beeld oproepen! Maar ook als wel alle drie categorieën zijn gegeven valt er nog genoeg te goochelen. Wat dacht u van de volgende uitspraak:

De groep docenten die voor doeleinden van determinatie meerkeuzevragen prefereert of geen voorkeur heeft, is in zijn totaliteit groter dan die van de voorstanders van open vragen. Dit resultaat is voor de leiding van het VAS onvoldoende aanleiding om de keuze voor uitsluitend meerkeuzevragen in de wiskundetoets te herzien.

De percentages zijn: van de ondervraagde docenten heeft 42% voorkeur voor open, 33% voor gesloten en 25% heeft geen voorkeur. Gezien alle detailopmerkingen van docenten over de individuele vragen en hun antwoord op deze laatste vraag, zou ik liever als conclusie formuleren:

Gezien de mening van docenten dat open vragen in het proces van determinatie veel meer inzicht geven in wat leerlingen kunnen en het feit dat slechts 33% van hen de voorkeur geeft aan meerkeuzevragen, komt de leiding van het VAS terug op de eerder genomen beslissing om de wiskundetoets uitsluitend in meerkeuzevorm aan te bieden.

Maar natuurlijk ben ik niet een instituut met commerciële belangen dat zijn product verdedigt.

In de artikelen wordt veelvuldig gebruikgemaakt van de

centrummaat 'gemiddelde'. Er wordt steeds beweerd dat gemiddeld over alle negen opgaven 65% van de docenten vindt dat open en gesloten vragen dezelfde wiskundige vaardigheid toetsen en dat slechts eenderde van de docenten dat niet vindt. Natuurlijk is zo'n uitspraak onzin: het bepalen van een gemiddelde van percentages bij verschillende typen vragen is getalsmatig wel mogelijk, maar dat getal zegt niets. Genuanceerde uitspraken van de docenten over verschillende typen vraagstellingen moeten elk afzonderlijk worden beoordeeld en geanalyseerd. Zo vindt steeds ruim 80% van de docenten dat meerkeuzevragen ('gesloten') en korte antwoordvragen, dat wil zeggen antwoorden zonder een getoonde berekening ('open'), dezelfde vaardigheid toetsen (opgaven 1, 5, 7 en 9 uit de lijst). Een minderheid van ongeveer 36% vindt dat dezelfde vaardigheden worden getoetst bij het al eerder genoemde exact berekenen in de open vorm en het kiezen uit intervallen bij de gesloten vorm (34% bij opgave 3 en 40% bij opgave 6). Het heeft geen enkele zin om deze zes percentages te middelen. Juist de uitersten die optreden zouden aanleiding moeten zijn voor de onderzoekers om te bezien of de opgaven waarbij deze verschillen optreden wel zo vergelijkbaar zijn.

Er wordt ook statistiek bedreven bij het onderzoek of gesloten en open vragen dezelfde wiskundige vaardigheden toetsen. Dat wordt gedaan door de twee versies van een toets voor te leggen aan verschillende groepen leerlingen, en dan te kijken of de toets *in zijn totaliteit* voor beide soorten groepen een hoge correlatie vertoont. Daarvoor bestaat allerlei geavanceerde programmatuur waarmee psychometrische gelijkwaardigheid en betrouwbaarheid kunnen worden becijferd. Om het onderzoek 'statistisch zuiver' te houden moeten de antwoorden wel op dezelfde manier worden nagekeken. Omdat bij meerkeuze alleen goed-fout mogelijk is, mogen de open vragen ook alleen maar gescoord worden met goed-fout (dat heet met een mooi woord dichotoom scoren). Het lijkt mij nogal wies dat je dan dezelfde resultaten krijgt. Gelukkig vinden de onderzoekers dan ook als resultaat 'dat niet kan worden vastgesteld dat de toetsen iets anders meten'. Maar ... vervolgens geven ze toch wel toe dat een totale toets niets zegt over een individuele vraag en dat polytoom gescoorde vragen betrouwbaarder zijn dan meerkeuzevragen. Kortom: hun onderzoek naar de gelijkwaardigheid van open en gesloten vragen proberen ze te beantwoorden door scores op een volledige toets te bekijken. Daarmee krijg je natuurlijk nooit een antwoord op je eigen onderzoeksvraag, laat staan dat je mag concluderen dat open en gesloten vragen dezelfde wiskundige vaardigheid toetsen.

Een opgave bewerkt

Door mijn lange betrokkenheid bij toetsontwikkeling kan ik het niet laten om bij een toetsvraag (open of gesloten) te kijken of het wel een goede vraag is. Het plaatje bij het veervraagstuk deugt niet. 'Lengte veer met gewicht' zou moeten slaan op het in het vraagstuk genoemde gewicht van 4 kg. Maar ik zie zo dat de veer met gewicht minder

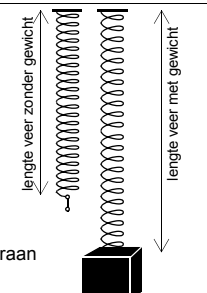
dan 30 cm lang is, want die andere (zonder gewicht) is 20 cm. Dat kan bij de gesloten vorm (onbedoelde) verwarring geven: het goede antwoord (32 cm) klopt niet met het plaatje, een fout antwoord (27 cm) wel. Dat brengt me op een andere hobby: proberen een vraag op een hoger niveau te krijgen door er wat aan te veranderen. Onderstaande versie van het veervraagstuk doet een beroep op veel meer vaardigheden: het doorzien van de structuur van een formule, meten en met schaal werken. Wat zou een docent daaruit veel kunnen aflezen over de vaardigheden van een leerling:

Als je aan een veer een gewicht hangt, wordt de veer langer. Hoe groter het gewicht, des te langer de veer wordt.

Met de formule hieronder kun je de lengte van de veer berekenen.

$$\text{lengte veer} = 20 + 3 \times \text{aantal kg}$$

De lengte van de veer is in centimeters. Het rechterplaatje toont de veer met daaraan een gewicht. Hoe groot is dat gewicht?



Ten slotte

Ik hoop dat duidelijk is gemaakt waarom het VAS geen goed instrument is voor determinatiedoeleinden. Verder hoop ik dat deze reactie discussie losmaakt over wiskundige vaardigheden en toetsvormen. Wij zijn in ieder geval van plan om meer te gaan schrijven over deze onderwerpen. Als laatste een mooi gedichtje (Wilmink, 1995):

Wat motten doen voor je ondergoed,
 wat de zure regen voor bomen doet,
 wat satan deed voor het Paradijs,
 doet multiple choice voor het onderwijs.

Henk van der Kooij, Freudenthal Instituut, Utrecht

Literatuur

- Dekker, T. & N. Querelle (2001). *The Great Assessment Picture book*. Te vinden op: www.fi.uu.nl/catch
- Kuhlemeier, H. & M. Steentjes (2002a). Docenten wiskunde over open en meerkeuzevragen: genuanceerder dan u denkt! *Nieuwe Wiskrant, Tijdschrift voor Nederlands Wiskundeonderwijs*, 22(2), 12-17.
- Kuhlemeier, H. & M. Steentjes (2002b). De vergelijkbaarheid van open en meerkeuzevragen bij wiskunde in het VMBO. *Nieuwe Wiskrant, Tijdschrift voor Nederlands Wiskundeonderwijs*, 22(3), 24-28.
- Lange, J. de (1987). *Mathematics, insight and meaning*. Utrecht: Freudenthal Instituut.
- Lange, J. de (1987). *Framework for Classroom Assessment in Mathematics*. Te vinden op: www.fi.uu.nl/catch
- OECD (2003). *The PISA 2003 Assessment Framework*. Parijs: OECD.
- Wilmink, W. (1995). Multiple choice. In: *Ernstig genoeg*. Amsterdam: Bakker.