

In het voorjaar van 2002 maakten ruim tweeduizend VMBO-leerlingen een wiskunde-toets zowel in de open als de meerkeuzevorm. De vraag was in hoeverre open en meerkeuzevragen vergelijkbaar zijn qua gemeten vaardigheden, moeilijkheid, betrouwbaarheid en afnametijd. In dit artikel doen **Hans Kuhlemeier** en **Melanie Steentjes** verslag van het onderzoek en de resultaten.

De vergelijkbaarheid van open en meerkeuzevragen bij wiskunde in het VMBO

Inleiding

Volgens velen doen open vragen een beroep op andere vaardigheden dan meerkeuzevragen. Meerkeuzevragen zouden ‘slechts’ passieve herkenning of reproductie vereisen, terwijl open vragen een beroep zouden doen op ‘hogere’ vaardigheden die samenhangen met het actief produceren van informatie. Hoe denken wiskundedocenten nu over deze kwestie? In het decembernummer van dit tijdschrift rapporteerden wij over een onderzoek waarin wij wiskundedocenten negen opgaven zowel in de open als in de meerkeuzevorm voorlegden (Kuhlemeier & Steentjes, 2002). Op de vraag of beide vraagvormen dezelfde vaardigheden meten, bleken wiskundedocenten verrassend genuanceerde antwoorden te geven. Het percentage docenten dat geen verschil zag, varieerde afhankelijk van de opgave van 34% tot 84%. Gemiddeld over de negen opgaven was eenderde van de docenten van mening dat beide vraagvormen een beroep doen op verschillende wiskundige vaardigheden, terwijl tweederde de omgekeerde mening was toegedaan. Vaardigheden die volgens wiskundedocenten vooral bij de meerkeuzeversie een rol spelen, zijn het globaal of schattend rekenen, het raden van het goede antwoord en het terugrekenen vanuit de alternatieven naar het goede antwoord. Vooral van belang voor de open versies zijn het opschrijven van de onderliggende berekening of formulering, het exact uitrekenen en afronden van het antwoord en het geven van het antwoord in de juiste eenheid. In dit artikel nemen we de proef op de som. We brengen verslag uit van onderzoek naar het effect van de vraagvorm op de wiskundeprestaties van tweedeklassers in het VMBO.

Wat wilden we weten?

In dit onderzoek zijn we nagegaan of een wiskundetoets in de open vraagvorm een beroep doet op dezelfde wiskundige vaardigheden als een volledig vergelijkbare toets in de meerkeuzevorm. Waarom wilden we dit weten? Het onderzoek is uitgevoerd in het kader van het Volg- en adviesstelsel (VAS) van het Cito (Van Krimpen & Kuhlemeier, 2002). Deze toetsbatterij ondersteunt scholen bij

het plaatsen van leerlingen in een passende leerweg of opleidingstype. Een belangrijke toets uit het VAS is de wiskundetoets. Deze toets bestond oorspronkelijk geheel uit open vragen. Dit leidde tot klachten van schoolleiders over de bewerkelijkheid en tijdrovendheid van het nakijken van de open vragen. Naar aanleiding hiervan is besloten om de wiskundetoets met ingang van het schooljaar 2001-2002 uitsluitend uit meerkeuzevragen te laten bestaan. Het onderzoek heeft tot doel na te gaan in hoeverre deze beslissing terecht is geweest. We zijn nagegaan of open en meerkeuzevragen een beroep doen op dezelfde wiskundige vaardigheden. De vraag naar de gemeten vaardigheden is echter niet het enige vergelijkingscriterium. Daarnaast zijn er nog drie andere onderzoeksvragen onderzocht, die we hieronder toelichten.

De antwoorden op open vragen kan men op verschillende manieren nakijken. Twee bekende mogelijkheden zijn goed-fout scoring (dichotoom) en polytome scoring, waarbij ook gedeeltelijk goede antwoorden punten opleveren. Volgens velen geeft polytome scoring van open opgaven een rijkere beschrijving van de oplossingsprocedures en strategieën die leerlingen toepassen dan dichotome scoring. Gegeven dezelfde opgaven zouden bij polytome scoring dan ook andere vaardigheden in het geding zijn dan bij dichotome scoring. Onze tweede onderzoeksvraag was dan ook of de twee scoringswijzen dezelfde wiskundige vaardigheden van de leerlingen bestrijken.

Meerkeuzevragen zijn vaak makkelijker dan open vragen. Leerlingen met onvolledige kennis of een gebrekkige motivatie kunnen het goede antwoord raden. Ook kunnen zij profiteren van een vergelijk-en-streep-weg strategie, waarbij minder plausibele alternatieven successievelijk worden geëlimineerd. Verder kunnen meerkeuzevragen leerlingen een zogeheten terugverwijzing geven. Vindt de leerling het antwoord niet terug onder de alternatieven, dan zal hij of zij de opgave nogmaals proberen, net zolang tot een antwoord gevonden wordt dat overeenkomt met een van de alternatieven. Onze derde onderzoeksvraag was dan ook of wiskundevragen in de meerkeuzevorm inderdaad makkelijker zijn dan diezelfde vragen in de open

Achtergrondinformatie bij de onderzoeksvragen

De eerste vraag naar de gemeten vaardigheden is het allerbelangrijkst (Traub, 1993). Als open en meerkeuzevragen een beroep zouden doen op verschillende vaardigheden, is vergelijking van moeilijkheid, betrouwbaarheid en afnametijd niet zo zinvol. Waarom dit zo is, lichten we toe aan de hand van een voorbeeld. Stel de ene leerling maakt het wiskunde-examen in de open vraagvorm en de ander krijgt diezelfde examenopgaven in de meerkeuzevorm voorgelegd. Ook al behalen beide leerlingen precies evenveel punten, mag men hun prestatie dan zonder meer vergelijken? Wij hopen dat u zegt: 'Nee, natuurlijk niet, het examen kan in de ene vraagvorm namelijk iets anders meten dan in de andere. Dat zou toch neerkomen op het vergelijken van appels met peren'.

Hoe stellen we nu vast of open en meerkeuzetoetsen dezelfde vaardigheden meten? Daarvoor zijn statistische procedures ontwikkeld. Toetsversies worden als psychometrisch gelijkwaardig beschouwd als de correlatie tussen de toetsscores één is of de waarde van één dicht benadert. We mogen er dan vanuit gaan dat de toetsversies hetzelfde meten. Voor de positie die de leerlingen binnen de getoetste groep innemen maakt het dan weinig uit of zij de toets in de open of de meerkeuzeversie maakten.

Prestaties op wiskundetoetsen zijn niet alleen afhankelijk van de vraagvorm, maar ook van de manier van nakijken. Als we open en meerkeuzevragen met elkaar willen vergelijken, moeten we de antwoorden van de leerlingen op precies dezelfde wijze nakijken. Als de meerkeuzevragen dichotoom (goed-fout) gescoord worden, moet dat bij de open vragen ook zo gebeuren. Anders is het effect van de vraagvorm niet te onderscheiden van dat van de scoringswijze. Vandaar dat we de open vragen niet alleen polytoom, maar ook dichotoom gescoord hebben.

Bij de betrouwbaarheid gaat het erom hoe nauwkeurig een toets de kennis en vaardigheden van de leerlingen in kaart brengt. De ervaring leert dat toetsen nooit volledig betrouwbaar zijn. Een toetsscore bestaat altijd voor een deel uit meetfouten. Hoe kleiner het aandeel van de meetfouten, hoe hoger de betrouwbaarheid. De betrouwbaarheid wordt doorgaans uitgedrukt in een getal tussen nul en één, waarbij een hogere waarde indicatief is voor een hogere betrouwbaarheid. Vanwege onbetrouwbaarheid zullen twee toetsversies de leerlingen nooit helemaal op dezelfde wijze rangschikken, ook al meten ze in werkelijkheid precies hetzelfde. Vandaar dat er statistische procedures zijn ontwikkeld om de 'ware' correlatie tussen de toetsversies uit te rekenen na correctie voor deze onbetrouwbaarheid.

Meerkeuzevragen zijn zoals gezegd vaak makkelijker dan open vragen. Gegeven gelijke vaardigheden en een gelijke betrouwbaarheid, is een verschil in moeilijkheid niet onoverkomelijk. Zolang de toetsversies de leerlingen maar op gelijke wijze rangschikken, kan men de ruwe scores via een eenvoudige lineaire transformatie tot elkaar herleiden. Deze transformatie wordt ook wel equivaleren genoemd. Geen enkele transformatie kan toetsen echter vergelijkbaar maken als er verschillende vaardigheden in het geding zijn.

vraagvorm.

Bij de betrouwbaarheid gaat het erom hoe nauwkeurig een toets de kennis en vaardigheden van de leerlingen in kaart brengt. De vraag naar de toetsbetrouwbaarheid is van praktisch belang voor de onderwijspraktijk. Hoe hoger namelijk de betrouwbaarheid, hoe minder opgaven de leerlingen hoeven te maken en hoe meer tijd er overblijft voor waardevolle instructie en oefening. Als vierde onderzoeksvraag hebben we daarom onderzocht of de vraagvorm invloed heeft op de betrouwbaarheid en de afnametijd. Daarnaast zijn we nagegaan of polytome scoring betrouwbaarder is dan dichotome scoring.

Welk toetsmateriaal hebben we gebruikt?

In het onderzoek is de wiskundetoets uit het Volg- en adviessysteem (VAS) van het Cito als uitgangspunt genomen. Deze toets bestaat uit 41 open vragen die zijn afgestemd op wat tweedeklassers in de gemengde en theoretische leerweg van het VMBO aankunnen. Uit deze toets hebben wij een aselechte steekproef van achttien opgaven getrokken voor afname in één lesuur. De achttien opgaven geven een redelijke dekking van de wiskunde in de onderbouw van het VMBO. Van de achttien opgaven gaan er zes over het vakonderdeel rekenen, zes over algebra en zes over meetkunde. Negen van deze opgaven zijn ook gebruikt om wiskundedocenten te bevragen naar hun mening over de vergelijkbaarheid van open en meerkeuzevragen (Kuhlemeier & Steentjes, 2002).

Bij elk van de achttien open opgaven hebben wij twee meerkeuzeversies gemaakt. Een gewone meerkeuzevraag en diezelfde meerkeuzevraag, met als extra afleider 'Het goede antwoord staat er niet bij'. De meerkeuzeversies zijn geconstrueerd op basis van de gegevens uit een vooronderzoek. Van ongeveer honderd leerlingen is steeds bekeken welke antwoorden het meest frequent voorkwamen en hoe de verschillende (foute) antwoorden verklaard zouden kunnen worden. Er is geprobeerd om al deze foute antwoorden in de afleiders te verwerken. Waar dit niet mogelijk was, zijn de meest voorkomende foute antwoorden gekozen.

Om het effect van de vraagvorm zuiver vast te stellen, hebben wij ervoor gezorgd dat de versies van een opgave stamequivalent zijn. Dit wil zeggen dat de inleiding op de vraag- of probleemstelling en de stam van de opgave voor elke versie gelijk is. We hebben drie toetsboekjes met achttien opgaven samengesteld die alleen verschillen ten aanzien van de vraagvorm. Het eerste toetsboekje bevat alleen open vragen (O), het tweede alleen meerkeuzevragen met de extra afleider 'Het goede antwoord staat er niet bij' (G) en het derde alleen de gewone meerkeuzevragen (M). Figuur 1 toont de O- en G-versie van een opgave waarbij de leerling een getal moet invullen in een woordformule.

Open versie

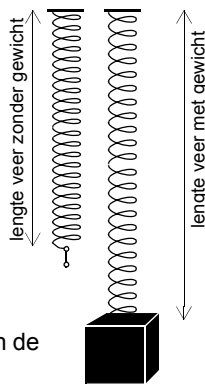
Als je aan een veer een gewicht hangt, wordt de veer langer.
Hoe groter het gewicht, des te langer de veer wordt.

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.

Je hangt een gewicht van 4 kg aan de veer.
Hoe lang is de veer nu?



Gesloten versie

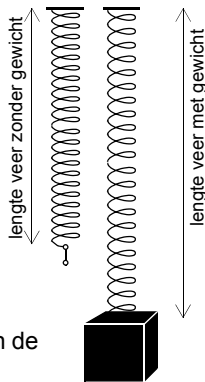
Als je aan een veer een gewicht hangt, wordt de veer langer.
Hoe groter het gewicht, des te langer de veer wordt.

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.

Je hangt een gewicht van 4 kg aan de veer.
Hoe lang is de veer nu?



- A. 27 cm
- B. 32 cm
- C. 92 cm
- D. Het goede antwoord staat er niet bij

fig. 1 Voorbeeld van een open opgave en een meerkeuze-opgave met als extra afleider 'Het goede antwoord staat er niet bij'

Verschillende vraagvormen en scoringswijzen: andere vaardigheden?

In ons onderzoek is nagegaan of open vragen (O), standaard meerkeuzevragen (M) en diezelfde meerkeuzevragen met als extra afleider 'Het goede antwoord staat er

Hoe zat het onderzoek in elkaar?

Aan het onderzoek is deelgenomen door ruim tweeduizend tweedeklassers uit de gemengde en theoretische leerweg van het VMBO. De leerlingen waren afkomstig van 104 scholen.

Om eventuele verschillen ten gevolge van de vraagvorm zo zuiver mogelijk vast te stellen, zijn de leerlingen twee keer getoetst. Daarbij maakten de leerlingen dezelfde opgaven, of in dezelfde vraagvorm, of in een verschillende vraagvorm. Er is voor gezorgd dat elke combinatie van vraagvormen voorkwam, dus ook twee keer dezelfde vraagvorm. De leerlingen zijn binnen schoolklassen systematisch aan deze negen combinaties i.c. groepen toegewezen (vergelijk Kuhlemeier, Steentjes & Kleintjes, in druk). Tussen de eerste en tweede afname zat een periode van twee à drie weken. Om het effect van herinnering en teruglopende motivatie zoveel mogelijk uit te schakelen, is ervoor gezorgd dat geen enkele leerling dezelfde vragen nogmaals in dezelfde volgorde moest maken.

De afnametijd is als volgt vastgesteld. Aan het begin van de afname gaf de docent de leerlingen de opdracht om de begintijd in het toetsboekje te noteren. De leerlingen leverden het toetsboekje bij de docent in zodra zij de laatste vraag beantwoord hadden. De docent noteerde de eindtijd op de daarvoor bestemde plaats in het toetsboekje.

niet bij' (G) een beroep doen op dezelfde wiskundige vaardigheden. Het zou te ver voeren om de uitgevoerde analyses en uitkomsten in detail te bespreken. Geïnteresseerde lezers worden daartoe verwezen naar het hoofd-rapport (Kuhlemeier, Steentjes & Kleintjes, in druk). We volstaan hier met te vermelden dat we het veronderstelde verschil in gemeten vaardigheden niet hebben kunnen aantonen. Van drie wiskundetoetsen, die met uitzondering van de vraagvorm qua vakinhoud en scoringswijze volledig identiek zijn, kan niet worden vastgesteld dat zij iets anders meten. Na correctie voor onbetrouwbaarheid wijken de correlaties tussen de O-, M- en G-toetsen niet significant af van één. De drie toetsen rangschikken de leerlingen dus op vrijwel dezelfde wijze. Deze bevinding komt overeen met Traub (1993). Hij komt op basis van enkele methodologisch zuivere studies tot de voorzichtige conclusie dat open en meerkeuzevragen in het kwantitatieve domein een beroep doen op dezelfde vaardigheden. Over een verklaring voor de afwezigheid van een effect van de vraagvorm kan op grond van ons onderzoek geen uitspraak worden gedaan. Het is echter verleidelijk deze verklaring te zoeken in de aard van de onderzochte wiskundeopgaven. De door ons gebruikte opgaven zijn zogeheten constructieopgaven. Bij dit type opgave is het antwoord niet direct vanuit het geheugen beschikbaar. De leerling kan niet volstaan met het herkennen van het goede antwoord en het antwoord is niet zonder meer in de opgave terug te vinden. De leerling moet zelf een oplossing voor een wiskundig probleem produceren via bijvoorbeeld een redenering of een berekening of via het verrichten van een meting met een geodriehoek of windroos.

Een aannemelijke verklaring voor de bevinding dat open en meerkeuzetoetsen hetzelfde meten, is dat de leerling deze mentale constructie *altijd* moet uitvoeren, ook al is de opgave in de meerkeuzevorm gesteld (vergelijk Traub, 1993).

Zoals gezegd, komt uit ons onderzoek bij leerlingen naar voren dat wiskundetoetsen die alleen verschillen voor wat betreft het vraagtype, hetzelfde meten. Deze hoofdconclusie staat ogenschijnlijk op gespannen voet met de mening van veel wiskundedocenten (zie de inleiding en onze bijdrage in het decembernummer 2002 van de *Nieuwe Wiskrant*). Een mogelijke verklaring verwijst naar het verschil tussen een toets en een individuele opgave. In het leerlingonderzoek zijn de geaggregeerde scores op drie versies van een volledige toets met elkaar vergeleken, terwijl de docenten hun mening gaven over de vergelijkbaarheid van drie versies van een individuele opgave. Dat de drie toetsversies maar weinig verschillen in moeilijkheid en hetzelfde meten, betekent niet automatisch dat dit ook geldt voor de open en meerkeuzeversies van elke individuele opgave afzonderlijk (Bennett, Rock & Wang, 1991). Bridgeman (1992) toonde overtuigend aan dat de somscores op toetsen met open en meerkeuzevragen vergelijkbaar kunnen zijn, ondanks grote verschillen op het niveau van de individuele opgaven in moeilijkheid en discriminerend vermogen. Met enige nadruk willen we er dan ook op wijzen dat de hoofdconclusies van ons onderzoek niet zonder meer overdraagbaar zijn naar de individuele opgaven. Er is hooguit aangetoond dat de achttien O-vragen in hun totaliteit een beroep doen op hetzelfde complex aan wiskundevaardigheden als de achttien G-vragen en de achttien M-vragen. Dit wil uiteraard niet zeggen dat elke individuele opgave in de ene vraagvorm precies hetzelfde meet als in de andere.

Velen veronderstellen dat polytome scoring van open opgaven een rijkere beschrijving van de oplossingsprocedures en strategieën van leerlingen geeft dan dichotome scoring. Met de door ons gebruikte opgaven en analyse-technieken hebben we niet kunnen aantonen dat de ene scoringswijze een beroep doet op andere wiskundige vaardigheden dan de andere (vergelijk Kuhlemeier, Steentjes & Kleintjes, in druk). Bij de interpretatie moet men echter bedenken dat de beantwoording van onze opgaven een beperkt aantal denkstappen vereist. De speelruimte om de polytome scoring te variëren al naar gelang de volledigheid of rijkdom van het oplossingsproces is met andere woorden beperkt. Complexere opgaven en uitgebreidere antwoorden bieden meer mogelijkheden om verschillende vaardigheidsaspecten in de scoring van de open vragen te betrekken. Zo zou men punten kunnen toekennen voor de correctheid van de berekening, de adequaatheid van de redenering en de helderheid van de communicatie. Wellicht hadden wij dan wel duidelijke verschillen gevonden tussen dichotome en polytome scoring.

Moeilijkheid

Naarmate de vraagvorm meer open is, zullen leerlingen minder profiteren van de raadkans, minder baat hebben bij terugverwijzing en hun antwoord meer zelf moeten produceren en formuleren. Wij verwachtten dan ook dat O-vragen gemiddeld moeilijker zouden zijn dan G-vragen, die op hun beurt weer moeilijker zouden zijn dan M-vragen. Dit blijkt inderdaad het geval, al is het verschil niet erg groot. De dichotoom gescoorde O-toets (68% goed) blijkt inderdaad het moeilijkst, gevolgd door de G-toets (72%) en de M-toets (75%). Zoals verwacht, blijkt de moeilijkheid van de G-vraag dus het midden te houden tussen de O- en de M-vraag.

Betrouwbaarheid en afnametijd

Bij meerkeuzevragen kunnen minder vaardige leerlingen het goede antwoord raden. Raden is niet alleen vanuit didactisch oogpunt ongewenst, het kan ook de toetsbetrouwbaarheid aantasten, omdat het toeval een grotere kans krijgt. Naarmate de vraagvorm meer open is, zal raden minder kans krijgen. Onze voorspelling was dan ook dat O-vragen betrouwbaarder zouden zijn dan G-vragen, die op hun beurt weer betrouwbaarder zouden zijn dan M-vragen. In ons onderzoek blijkt dit niet het geval. De betrouwbaarheid waarmee de achttien O-, G- en M-vragen de wiskundige vaardigheid meten, bedraagt respectievelijk .60, .64 en .67. De M-vragen zijn dus betrouwbaarder dan de G-vragen, die op hun beurt weer betrouwbaarder zijn dan dichotoom gescoorde O-vragen. Ook ten aanzien van de betrouwbaarheid neemt de G-vraag dus een middenpositie in tussen de beide andere vraagvormen.

Polytome scoring van open vragen blijkt aanzienlijk betrouwbaarder dan dichotome scoring (.72 versus .61). Open vragen lijken dus niet zozeer betrouwbaarder vanwege de vraagvorm an sich, maar meer vanwege de wat genuanceerdere scoring. Bij de interpretatie moeten we bedenken dat slechts zes van de achttien opgaven polytoom gescoord zijn. Hadden we alle opgaven polytoom kunnen scoren, dan was het verschil mogelijk groter geweest.

Bij open vragen moeten de leerlingen hun antwoorden in leesbare en begrijpelijke taal opschrijven en daarnaast vaak ook nog eens toelichten. Wij verwachtten dan ook dat een open vraag gemiddeld meer afnametijd zou vergen dan een meerkeuzevraag. Van een G-opgave bieden de alternatieven de leerling minder steun bij het vinden van het juiste antwoord. Daardoor zouden G-vragen meer afnametijd vergen dan standaard meerkeuzevragen. Echter, de afnametijd zal korter zijn dan bij O-vragen, aangezien de leerling het antwoord niet in leesbare en begrijpelijke taal hoeft op te schrijven. In hoeverre vinden deze veronderstellingen ondersteuning in de gegevens? Met het beantwoorden van de achttien open vragen zijn de leerlingen gemiddeld vier minuten langer bezig dan met

de meerkeuzevragen (respectievelijk 21 versus 17 minuten). Van een gemiddelde open en meerkeuzevraag is de afnametijd 71 respectievelijk 56 seconden. Een verschil in afnametijd tussen de beide typen meerkeuzevragen kon niet worden vastgesteld.

Polytome gescoorde open vragen zijn zoals we hiervoor zagen betrouwbaarder dan meerkeuzevragen, maar daar staat tegenover dat de beantwoording ervan meer tijd kost. Een legitieme vraag is dan ook of open vragen nog steeds betrouwbaarder zijn als men rekening houdt met het feit dat men er in dezelfde tijd minder van kan afnemen. Onder het constant houden van de afnametijd repteert er nog steeds een verschil in het voordeel van de polytome gescoorde open opgaven (.72 versus .66).

Reikwijdte van de bevindingen

De uitkomsten van dit onderzoek moeten we interpreteren in het licht van de wijze waarop de gegevens verzameld zijn. Een eerste kanttekening betreft de moeilijkheidsgraad van de gebruikte opgaven in relatie tot de onderzochte groep leerlingen. Het effect van de drie vraagvormen is onderzocht aan de hand van achttien constructie-opgaven afkomstig uit de wiskundetoets van het Volg- en adviessysteem. De moeilijkheidsgraad van deze opgaven is goed afgestemd op het vaardigheidsniveau van leerlingen in het tweede leerjaar van de gemengde en theoretische leerweg van het VMBO. Hadden wij deze leerlingen bijvoorbeeld een set veel te moeilijke opgaven voorgelegd, dan waren de conclusies wellicht anders geweest, bijvoorbeeld omdat effecten van raden en terugverwijzing zich vooral zullen voordoen als leerlingen met (te) moeilijke opgaven geconfronteerd worden.

Een tweede kanttekening betreft de gehanteerde scoringswijzen. Behalve op dichotome wijze zijn de open opgaven ook nagekeken met het gedeeltelijk polytome scoringsvoorschrift van de wiskundetoets uit het Volg- en adviessysteem. De resultaten maken aannemelijk dat polytome scoring een beroep doet op dezelfde wiskundige vaardigheden als dichotome scoring. Bij de interpretatie moet men echter bedenken dat de beantwoording van onze opgaven een beperkt aantal denkstappen vereist. De speelruimte om de polytome scoring te variëren al naar gelang de volledigheid of rijkdom van het oplossingsproces is met andere woorden beperkt. Uitgebreidere antwoorden bieden meer mogelijkheden om verschillende vaardigheidsaspecten in de scoring van de open vragen te betrekken. Zo zou men punten kunnen toekennen voor de correctheid van de berekening, de adequaatheid van de redenering en de helderheid van de communicatie. Wellicht hadden wij dan wel een verschil gevonden tussen dichotome en polytome scoring.

Een derde kanttekening betreft het doel van de toetsing. Het onderzoek is uitgevoerd in de context van het Volg- en adviessysteem. Het doel van deze toetsbatterij is het volgen van leerlingen teneinde hen in een geschikte leerweg of opleidingstype te plaatsen. Ons onderzoek laat

zien dat open en meerkeuzevragen uitwisselbaar zijn in de zin dat zij een beroep doen op dezelfde wiskundige vaardigheden. De veronderstelde verschillen in moeilijkheid en betrouwbaarheid bleken of niet aanwezig of vrij klein. Dit neemt uiteraard niet weg dat de vraagvorm voor andere toetsdoelen wel van belang kan zijn. Wiskundeleraars gebruiken toetsen ook voor voortgangscntrole. Dat wil zeggen: het opsporen van tekorten teneinde deze vervolgens te kunnen wegwerken. Open vragen bieden de wiskundeleraar meestal meer mogelijkheden om fouten in het oplossingsproces te diagnosticeren dan meerkeuzevragen. Voor doeleinden van voortgangscntrole zijn open en meerkeuzevragen uiteraard niet uitwisselbaar.

Aanbeveling

De praktische aanleiding tot het onderzoek is gelegen in klachten van schoolleiders over de bewerkelijkheid van het nakijken van de open wiskundevragen uit het Volg- en adviessysteem. Om de werklust voor de docenten te beperken en de scholen sneller te kunnen informeren over de toetsresultaten, bevatten de wiskundetoetsen in het VAS tegenwoordig uitsluitend meerkeuzevragen. De uitkomsten van het onderzoek geven onvoldoende aanleiding om de keuze voor meerkeuzevragen te heroverwegen.

Hans Kuhlemeier en Melanie Steentjes, Cito, Arnhem

Literatuur

- Bennett, R.E., D.A. Rock & M. Wang (1991). Equivalence of free-response and multiple-choice-items. *Journal of Educational Measurement*, 28(1), 77-92.
- Bridgeman, B. (1992). A comparison of quantitative questions in open-ended and multiple-choice formats. *Journal of Educational Measurement*, 29(3), 253-271.
- Krimpen, W. van & H. Kuhlemeier (2002). Alle stukjes van de puzzel. Determineren en adviseren in de onderbouw van het voortgezet onderwijs. *Handboek Studiehuis Basisvorming*, 8.22, 1-10.
- Kuhlemeier, H. & M. Steentjes (2002). Docenten wiskunde over open en meerkeuzevragen: verrassend genuanceerder dan u denkt! *Nieuwe Wiskrant*, 22(2), 12-17.
- Kuhlemeier, H., M. Steentjes & F. Kleintjes (in druk). *De gelijkwaardigheid van open en meerkeuzevragen bij wiskunde. Effecten van vraagtype en scoringswijze op gemeten vaardigheden, betrouwbaarheid, moeilijkheid en afnametijd*. Arnhem: Citogroep.
- Traub, R. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. In R. Bennett & W. Ward (eds.). *Construction versus choice in cognitive measurement* (pp. 29-44). Hillsdale, NJ: Lawrence Erlbaum Associates.