

Universiteit van Amsterdam
cowo Centrum voor Onderzoek van het Wetenschappelijk
Onderwijs

Voetboogstraat 1 / Postadres: Spui 21 1 Tel.: 525 2835

De keuze tussen meerkeuze vragen en openeind vragen voor
de
toets.

Ben Wilbrink. 5-3-1976

Voor beoordeling van opgedane kennis en inzichten staan de docent een groot aantal verschillende vraagvormen ter beschikking, uiteenlopend van de mondelinge ondervraging tot de (eventueel landelijk gestandaardiseerde) meerkeuze studietoets. Welke vraagvorm voor welke leerstof het meest geschikt is, hangt van de aard van die leerinhoud af, van de bedoeling van de toetsing, het aantal te beoordelen studenten, etcetera. Hetzelfde geldt voor de veel beperktere keuze waarvoor studenten en docenten zich vaak gesteld zien, namelijk of in een bepaalde situatie de voorkeur uit moet gaan naar de meerkeuze vraagvorm, of naar een vraagvorm waarbij korte antwoorden door de student zelf geconstrueerd moeten worden. Deze keuzeproblematiek is het onderwerp van dit advies, en de korte literatuurbespreking die als bijlage te verkrijgen is. Aangenomen wordt dat wat toetsingsvorm betreft de keuze voor korte schriftelijke vragen adequaat is n.b.t. de aard van de leerinhoud en de bedoeling van de toetsing.

Een eerste, oppervlakkige, bestudering van de literatuur op dit punt (De Groot & Van Naerssen 1973, Thorndike 1971, Stanley & Hopkins 1972 onder andere) maakt al snel duidelijk dat aan de vergelijking van de kortantwoord vragen (aanvul- of openeind vragen e.d.) met meerkeuze vragen noch empirisch, noch theoretisch aandacht is besteed. Jawel, er is veel onderzoek gedaan naar het verschil in betrouwbaarheid van beoordeling tussen opstellen en meerkeuze toetsen (Mellenbergh, in De Groot en Van Naerssen 1973), maar dergelijke vergelijkingen zijn in twee opzichten voor ons keuzeprobleem irrelevant: Ten eerste is onze vraag niet of het opstel, dan wel de meerkeuze toets gebruikt moet

worden, maar is de vraag of de toets moet bestaan uit vragen met geprecodeerde antwoorden (meerkeuze), of uit een groot (zeg evenveel als bij een meerkeuze toets) aantal vragen die een kort (een woord, een zin) antwoord vragen dat de student zélf moet formuleren.

Ten tweede zijn verschillen in beoordelaar betrouwbaarheid tussen de ene vorm van toetsen en de andere vorm van toetsen slechts van ondergeschikt belang: tegen geringere betrouwbaarheid kun je je op verschillende manieren wapenen, waarvan de eenvoudigste manier wel is de toets t'e verlengen.

Argumenten die moeten pleiten voor de betere kwaliteit van de meerkeuze toetsvorm blijken, bedoeld of onbedoeld, vrijwel altijd te berusten op het zwaar overtrokken en gedeeltelijk ook onterechte voordeel van de objectieve scorbaarheid van meerkeuze toetsen, en daarnaast op de oneerlijke vergelijking die telkens weer gemaakt wordt tussen de meerkeuze toets die op uiterst zorgvuldige wijze door samenwerking van vakdeskundigen, itemconstructeurs en psychometrici vervaardigd zijn met opstel of openeind vraagvormen die in een vloek en een zucht door de docent zelf, zonder deskundige ondersteuning en de mogelijkheden van uitgebreide tryouts, in elkaar gezet worden. Bij de bespreking van de literatuur in de afzonderlijk aan te vragen bijlage blijkt dit inderdaad telkens weer het geval te zijn.

De docent in het hoger onderwijs zal meestal zonder deskundige ondersteuning zijn toets moeten construeren, of dat nu een toets wordt die bestaat uit meerkeuze vragen, of een toets die bestaat uit een groot aantal vragen waarop de student een kort antwoord moet formuleren. Dat betekent onmiddellijk dat alle voordelen die aan de meerkeuze toets worden toegeschreven, en die slechts realiseerbaar zijn wanneer de toets door item constructeurs, vakdeskundigen, en psychometrici in samenwerking in elkaar gezet en uitgeprobeerd wordt, niet relevant zijn voor de docent die moet kiezen tussen deze vraagvorm en het meest voor de hand liggende alternatief: de openeind vraagvorm.

Dan is het vervolgens niet onwaarschijnlijk dat verdedigers en

gebruikers van de meerkeuze toets een aantal verborgen motieven hebben (Hofstee 1973), die het onderwijs grote kwalitatieve en economische schade kunnen berokkenen, om nog maar te zwijgen van de onnodige stress waarmee de student in dergelijke onderwijssituaties geconfronteerd wordt.

Wanneer gebruik je meerkeuze vragen', wanneer open vragen?

Er zijn een paar veel voorkomende onderwijssituaties waarvoor deze vraag zich zonder meer duidelijk laat beantwoorden.

Waar antwoorden berekend of geconstrueerd moeten worden, of waar met kaarten en diagrammen gewerkt kan worden, moet men bij voorkeur werken met vragen waarbij de student een kort (een enkele regel, een enkel woord of getal) moet formuleren; open vragen dus.

Bij gevraagde berekeningen of constructies de student laten kiezen uit een aantal gegeven antwoordmogelijkheden is een beetje pervers, introduceert interpretatie en betrouwbaarheidsproblemen (raadkans!), en is eigenlijk alleen maar goed te praten in situaties waarin geen mogelijkheid bestaat om het werk snel en efficiënt nagekeken te krijgen anders dan door van computer scoring gebruik te maken.

Vragen die betrekking hebben op kaarten en diagrammen zijn eigenlijk qua karakter al enigszins van het meerkeuze type: alle alternatieven zijn a.h.w. in de vraag al gegeven, en de student kiest er uit (eventueel combineert daaruit).

Beide soorten stof lenen zich bij uitstek voor korte, eenduidige beantwoording, het soort beantwoording waarover achteraf weinig of geen problemen hoeven te rijzen of een gegeven antwoord nu fout of goed gerekend moet worden. Een enkel twijfelgeval zal zich altijd wel voordoen, evenals bij meerkeuze items de twijfels achteraf opkomen naar aanleiding van de vraagformulering zélf die kan blijken ook als onjuist gescoorde alternatieven niet volkomen

als onjuist uit te sluiten.

Wanneer het aantal studenten niet groot is, zeg niet meer dan 300 of 500 afhankelijk van de aard van stof en vraagstellingen, open. vragen gebruiken.

Een kwestie van efficiency: het is met kleinere aantallen studenten niet voordelig om de meerkeuze vorm te kiezen. Waar de grens precies getrokken moet worden, is niet in haar algemeenheid te beantwoorden. Dat kan er o.a. van af hangen of men de studenten zélf de toets kan laten nakijken; of door het vermijden van de raatkans problematiek het aantal vragen belangrijk gereduceerd kan worden zonder verlies in betrouwbaarheid; of de stof zich makkelijk leent tot vraagformuleringen die zich eenduidig laten beantwoorden; .e.d.

Eenvoudige kennis vragen lenen zich uiteraard ook bijzonder goed voor de openeind vorm. Het is echter in de meeste onderwijs situaties niet waarschijnlijk (althans, in het hoger onderwijs) dat kennisvragen een belangrijk onderdeel van de doelstellingen uitmaken.

Dan is er een vrij omvangrijke categorie leerinhouden die zich maar met moeite laten vangen in de vorm van vragen die zich kort en eenduidig laten beantwoorden. Voor deze stof geldt natuurlijk hetzelfde probleem waar het gaat om het zoeken naar geschikte formuleringen voor de meerkeuze vraagvorm, Ook hier blijkt op deze grond geen voorkeur voor meerkeuze vragen afgeleid te kunnen worden. Het schijnbare voordeel van de meerkeuze vraagvorm, in dit geval, is dat de antwoorden zich eenvoudig laten scoren. Daar staat tegenover dat naarmate het moeilijker is om een bevredigende vraagformulering te vinden, de wijze van beantwoording door de student steeds minder duidelijk wordt, ook al heeft men de alternatieven waaruit hij moet kiezen bij het meerkeuze item al geformuleerd. Het laatste betekent dat betrouwbaarheid en validiteit van deze vragen sterk terugloopt, een nadeel dat veel sterker weegt dan het voordeel dat een meerkeuze item tenminste objectief (d.w.z. door de computer uitvoerbaar) scorebaar is.

Wat men in genoemde gevallen kan doen, wanneer desnoods met enige moeite toch goede vragen met korte antwoorden geformuleerd kunnen worden, bij voorkeur gebruik maken van de open vraagvorm. Wanneer het om zeer grote aantallen studenten gaat, dan schakelt men bij voorkeur een team van vakdeskundigen, itemconstructeurs, en psychometrici in om de toets, of liever een item verzameling te construeren en uit te proberen. Bij niet grote aantallen studenten gebruikt men de open vraagvorm, maar dan hoort daar wel bij dat er maatregelen genomen worden om de beoordeling kwalitatief veilig te stellen.: ieder werk (bij voorkeur vragen afzonderlijk, of groepjes vragen afzonderlijk) door meerdere beoordelaars (eventueel uit de studenten te recruteren) na laten kijken aan de hand van beoordelingsvoorschriften.

Is het niet meer doenlijk om behoorlijke vragen met korte antwoorden te construeren, dan moet omgekeken worden naar andere toetsingsvormen, of de toetsfilosofie moet omgebogen worden op de wijze zoals door Hofstee gepropageerd (Hofstee 1973).

Enkele slotopmerkingen.

Een groot nadeel van de meerkeuze vraagvorm is dat de wijze waarop de constructie van deze vragen plaats vindt nog sterk afhankelijk is van intuities en invallen van de docent, en bestaande tradities. Er is met andere woorden.nog geen methodiek ontwikkeld voor het formuleren van onderwijskundig relevante (d.i. op de doelstellingen van het onderwijs afgestemde) meerkeuze vragen. Men zal op dit moment de literatuur nog vergeefs afzoeken naar dergelijke constructie regels. Wat wél in ieder handboek uitgebreid beschreven staat zijn de trucjes in de vraagformulering (los van de leerinhouden), en de fouten die men bij de formulering (ook weer: los van de leerinhouden) moet vermijden. Een en ander betekent dat meerkeuze vragen voor de student vaak telkens weer verrassingen inhouden, dat dingen gevraagd worden waarop hij zich niet heeft voorbereid, of had kunnen voorbereiden. Daarnaast wordt in de meerkeuze vraagvorm van de student vaak verlangd dat hij onderscheidingen kan maken tussen de

aangeboden alternatieven; onderscheidingen die vaak als zodanig in de leerstof, de studieboeken, in het geheel niet aan de orde zijn geweest; waar hij zich dus, wederom, vaak niet op geprepareerd heeft of op heeft kunnen prepareren. Beide bezwaren zullen zich juist sterk doen gelden bij door de docent zelf gemaakte toetsen. omdat de zorgvuldig door deskundigen ontwikkelde en uitgeteste toets minder aan dergelijke euvels zal lijden. Hoewel het de docent niet kwalijk te nemen is dat hij in deze opzichten geen goede toets kan construeren omdat zijn tijd en constructievaardigheid noodzakelijk beperkt zijn, zou hij toch wel terdege moeten overwegen of zijn keuze voor de meerkeuze vraagvorm wel te rechtvaardigen is tegenover het alternatief van de bekende, oeroude en vertrouwde open vraag vorm.

Een ander belangrijk bezwaar tegen de meerkeuze vorm, een bezwaar dat evenals de juist genoemde nauwelijks onderzocht lijkt te zijn en ook weinig of niet gesignaleerd wordt door auteurs op toetsconstructie-gebied, is dat de student in grote onzekerheid gelaten wordt of hij de juiste antwoorden heeft aangekruist of niet. Aan dit bezwaar kan gedeeltelijk tegemoet gekomen worden door de nog vrijwel niet in praktijk gebrachte mogelijkheid onmiddellijk na de toetsafname de scoring te verrichten en de studenten zowel de vragen als de juiste antwoorden te verstrekken. Er zijn natuurlijk weer slimme vogels die uit die onzekerheid van de student munt proberen te slaan door naast aankruisen van het juist geachte alternatief ook van de student te vragen aan te geven of hij al dan niet zeker is van het gegeven antwoord. (Zekerheidsscore, zie o.a. Sandbergen 1973, Kansup & Hakstian 1975, Hakstian & Kansup 1975) De scoring wordt dan aangepast, zodat een extra punt verkregen wordt wanneer bij het juist aangekruiste alternatief aangegeven was dat men er zeker van was, en een punt wordt afgetrokken wanneer bij een onjuist aangekruist alternatief aangegeven was dat men er zeker van was het juiste antwoord aangekruist te hebben. Hoewel Sandbergen zich gematigd optimistisch uitlaat over de voordelen van deze procedure is men er in het algemeen toch wat huiverig voor, en rechtvaardigen de onderzoekresultaten geen routinematige toepassing van deze techniek. Maar goed, het is ons nu niet om de techniek op zich te doen, maar om een illustratie van de mate waarin de student bij de meerkeuze toetsvorm in onzekerheid

gebracht en vaak gelaten wordt, een onzekerheid die groot genoeg is dat er meettechnisch munt uit geslagen kan worden. (dat de procedure desondanks niet geschikt is voor toepassing in het onderwijs komt omdat de 'winst' gekocht wordt tegen langer benodigde tijd voor het maken van de toets, en verlaging van de validiteit van de toets doordat verschillende studenten op verschillende wijze van deze zekerheidsaanduiding gebruik maken).

Een mogelijk nadeel van de openvraag vorm, althans wanneer het niet om reken of constructie opgaven en dergelijke gaat, is dat bij onzorgvuldige vraagstelling het gevaar bestaat dat de vraag ofwel pietepeuterige details uit een boek terugvraagt, details die ook de collega vakdeskundige niet zonder meer zal weten te geven. Maar tegen onzorgvuldige vraagformulering is, wanneer het kwaad geschied is, geen enkel kruid gewassen. Zowel de open vraagvorm als de meerkeuze vraagvorm vereisen de nodige zorgvuldigheid bij het construeren, waarbij de open vraagvorm in het voordeel is omdat daar niet naar plausibele afleiders gezocht hoeft te worden.

Een ander mogelijk nadeel van de open vraagvorm heeft zij gemeen met de meerkeuze items die in de vorm. van de ja-nee of juist-onjuist vraag gegoten worden: de vraag laat zich op meerdere manieren juist beantwoorden terwijl de docent die de vraag formuleerde juist één van die mogelijkheden als enig goede bedoelde. Ook hier is de remedie: zorgvuldige constructie van de vraag, en uitproberen door ze eerst aan collega's ter beantwoording voor te leggen en op basis van die resultaten de toets te 'zuiveren'.

Tenslotte de vraag of bij de open vraagvorm ook toets en itemanalyse met behulp van de computer uitgevoerd kunnen worden. Dat is eenvoudig mogelijk, door bij het nakijken van de open vragen een IBM scoringsformulier in te vullen. Een sterke vereenvoudiging zou daarbij nog bereikt kunnen worden door alleen foute antwoorden aan te strepen (in het eerste vakje horend bij de vraag), en als sleutel op te geven dat voor iedere vraag het tweede 'alternatief' het juiste is, er. dat 'overgeslagen vragen' ook juist zijn. Voor het laatste is waarschijnlijk een kleine wijziging in

het computerprogramma nodig, die echter eenvoudig uit te voeren is. Op deze wijze kan men ook voor de toets met open vragen gegevens krijgen zoals itemrest correlaties, KR 20 homogeniteit van de toets, standaard deviatie, en dergelijke. Wanneer echter het aantal vragen belangrijk kleiner is dan bij meerkeuze toetsen gebruikelijk, of wanneer de beantwoording van de vragen niet 'kort' meer is maar uitloopt in de richting van kleine opstellen, moeten andere technieken voor de analyse van de beoordelingsprocedure aangewend worden (Cronbach c.s. 1972; Hoekman & Neervoort 1975).

CONCLUSIE.

Over de vraag of de meerkeuze vorm te verkiezen is boven vraagvormen. die korte en eenduidige antwoorden vragen, is geen systematisch onderzoek gedaan. Wanneer de stof vraagt om het berekenen of construeren is dat onderzoek ook niet nodig: de open vraagvorm is hier duidelijk de meest geschikte, tenzij door de massaliteit van toetsafname efficiency overwegingen dwingen tot meerkeuze (bij landelijke toetsen bijv.). Beweerde voordelen van meerkeuze vraagvormen houden alleen stand bij vergelijking met essay beoordelingen (waar overigens ook betere technieken voor beschikbaar komen), en wanneer met zorg geconstrueerde meerkeuze tests vergeleken worden met haastig in elkaar gezette open vragen toetsen. Zolang echter aan het construeren van een openvragen toets dezelfde zorg besteed wordt als aan constructie van een meerkeuze toets, zullen in veel onderwijssituaties argumenten met betrekking tot de doorzichtigheid van de toetssituatie voor de student de keuze ten gunste van de openvraagvorm doen uitvallen.

Referenties.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. London: Wiley, 1972.

De Groot, A. D.,? & Van Naerssen, R. F. (red.) Studietoetsen,

construeren, afnemen, analyseren. Den Haag: Mouton, 1973

Hakstian, A. R., & Karsup, W. A comparison of several methods of assessing partial knowledge in multiple choice tests. *Journal of Educational Measurement*, 1975, 12, 231-240.

Hoekman, K., & Neervoort, Th. J. Generaliseerbaarheid van rating scores. In: *Onderwijsresearchdagen 1975*. Amsterdam: Werkgroep Onderwijs Research, p/a Onderwijsresearch Vrije Universiteit, De Boelelaan 1105 Amsterdam. 1975.

Hofstee, W. K. B. Participatiekontrolle door 'onbenullige' toetsitems. *Nederlands Tijdschrift voor de Psychologie*, 1973, 28, 189-198.

Karsup, W. & Hakstian, A. R. A comparison of several methods of assessing partial knowledge in multiple choice tests. *Journal of Educational Measurement*. 1975, 12, 219-230.

Sandbergen, S. Zekerheidsaanduiding bij het meten van studieprestaties: een studie naar de onderwijskundige en psychometrische aanvaardbaarheid van een toetsingstechniek. (Dissertatie) Amsterdam: Stichting Research Instituut voor de Toegepaste Psychologie, 1973,

Stanley, J. C., & Hopkins, K. D. *Educational and psychological measurement and evaluation*. Englewood Cliffs, N. J.: PrenticeHall, 1972.

Thorndike, R. L. (Editor) *Educational Measurement*. Washington, D. C.: American Council on Education, 1971.