

Uit de **praktijk**

□ Ab van Beek

Digitale innovatieve **vraagvormen:** verslag van een **studiemiddag**

Op een discussieplatform voor toetsen in het hoger onderwijs, de Toets HO-Lijst, ontstond in 2007 een heftige e-mail-discussie over het gebruik van innovatieve vraagvormen zoals de multiple response-vragen, ordenings-, sleep- en matching-vragen in digitale toetsen. De meningen waren verdeeld: sommigen vonden het gebruik van zulke vraagvormen ongewenst, anderen meenden dat deze vraagvormen nieuwe mogelijkheden konden bieden. De discussie was aanleiding voor het spontaan organiseren van een studiemiddag. Met een gemeenschappelijk doel – meer te weten komen om betere vragen en toetsen te kunnen ontwikkelen – zijn vijftig deelnemers bijeengekomen.

INNOVATIEVE VRAAGVORMEN

In feite zijn multiple responsevragen – vragen waarbij meer alternatieven goed kunnen zijn –, ordeningsvragen – vragen waarbij de alternatieven in de juiste volgorde moeten worden geplaatst –, sleep- en matchingvragen – vragen waarbij de bij elkaar passende gegevens moeten worden gezocht – geen echte innovatieve vraagvormen. Ze bestaan al veel langer maar worden vergeleken met de standaardmeerkeuzevragen relatief weinig gebruikt. Dit verandert door de komst van elektronische toetssystemen. Het gemak waarmee deze vraagvormen gemaakt en gescoord kunnen worden, heeft daaraan bijgedragen. Over het gebruik van deze vraagvormen in elektronische toetssystemen zijn betrekkelijk weinig onderzoeksgegevens bekend.

Aanleiding en organisatie

De Toets HO-lijst leidt een wat sluimerend bestaan, maar af en toe ontstaan korte discussies nadat een deelnemer een bericht heeft gestuurd naar alle leden van de lijst. In dit geval werd een opmerking geplaatst dat bij bepaalde vraagvormen, zoals de matchingvraag of de multiple responsevraag, altijd moet worden aangegeven hoeveel alternatieven juist zijn. Volgens Ineke Verheul van het Centrum Leermiddelenstudie te Utrecht (CLU) en degene die de e-maildiscussie op gang bracht, worden bij de multiple responsevraag zowel kennis als persoonskenmerken gemeten; bij de matchingvraag zou altijd sprake zijn van afhankelijkheid binnen de vraag. Enkele deelnemers van de Toets HO-lijst dachten hier anders over en er ontstond een levendige discussie. Tijdens de discussie werd het voorstel gedaan om een studiemiddag te organiseren over het onderwerp. Ineke Verheul, Onno Tomson (van de Question Mark gebruikersvereniging), Ab van Beek

Uit de praktijk

(secretaris van de TestVision gebruikersvereniging) en Silvester Draaijer (Vrije Universiteit) vormden vrijwillig de organisatie van de studiemiddag. Zij stelden een programma samen. Ondersteuning kwam van de kant van het Kenniscentrum ECABO en van de Nederlandse Vereniging voor Examens (NVE). Vijftig personen hebben de gratis toegankelijke studiemiddag bezocht.

Na een inleiding door drs. C.C. Verheul spraken achtereenvolgens dr. ir. H.J. Vos (Universiteit Twente), ir. S. Draaijer (VU) en van Teelen Kennismanagement ing. H.F.A.M. Molkenboer en mevrouw drs. J. Soeting. De bijdrage van Vos is in dit nummer van *EXAMENS* opgenomen.

Drie belangrijke discussiepunten

Volgens Verheul zou de discussie moeten gaan over de vragen:

- zijn er vraagvormen die je beter helemaal niet kunt gebruiken?
- Is het verstandig om binnen een toets (veel) verschillende vraagvormen te gebruiken?
- In hoeverre wordt bij het gebruik van nieuwe toetsvormen aan de eis van transparantie voldaan?

Ongewenste vraagvormen

Met name bij multiple response- en matchingvragen speelt de vraag hoe betrouwbaar deze vraagvormen zijn. Multiple responsevragen worden gebruikt als oplossing voor het probleem van 'rijtjesvragen'. Met traditionele meerkeuzevragen is het niet goed mogelijk om rijtjes van meer dan drie opties te bevragen. De oplossing om daarvoor multiple responsevragen te gebruiken, levert echter weer een ander probleem op wanneer niet wordt aangegeven hoeveel antwoorden er gegeven moeten worden. In dat geval worden ook persoonlijkheidskenmerken gemeten zoals de mate van zekerheid van de kandidaat. Bij matchingvragen leidt het kiezen van een onjuiste combinatie automatisch tot een volgende fout. Een manier om dit probleem op te lossen is om meer opties aan te bieden dan er nodig zijn, maar het is de vraag of daarmee niet weer andere mechanismen in het leven worden geroepen. Meet een dergelijke vraag nog wel wat iemand weet of meet zo'n vraag ook (en misschien zelfs vooral) de snelheid waarmee een kandidaat de juiste alternatieven uit een lange lijst weet te selecteren?

Verschillende vraagvormen in een toets

Standpunten voor het toepassen van verschillende vraagvormen binnen één toets variëren van 'doen, want dat werkt motiverend' tot 'liever niet, tenzij het niet anders kan'. Wat de motiverende werking van een toets betreft: Cronbach stelde (inderdaad, lang geleden!) dat een toets op zichzelf al motiverend genoeg werkt. Kandidaten weten dat ze beoordeeld worden en weten wat de consequenties

van zakken en slagen zijn. Alles wat aan een toets wordt toegevoegd om kandidaten 'leuk bezig te houden' werkt alleen maar storend. Variëren in werkvormen binnen het onderwijs: prima, maar niet in een toets! Niet alleen omdat een toets op zichzelf al motiverend genoeg is, maar ook omdat het storend is om steeds van de ene vraagvorm naar de andere te moeten overspringen. Is dat inmiddels een achterhaald standpunt? Aanbieders van digitale toetsen vinden van wel: zij krijgen geen klachten van kandidaten over de variatie aan vraagvormen. Gericht onderzoek naar het gebruik van verschillende vraagvormen door elkaar binnen één toets is er tot nu toe echter nog niet geweest.

Transparantie

Naast validiteit en betrouwbaarheid is transparantie – duidelijkheid over wat er van kandidaten verwacht wordt – een eis voor de kwaliteit van de toets. Transparantie over leerdoelen gecombineerd met een correcte samenhang tussen toets en leerdoelen zorgt ervoor dat het leerproces van een kandidaat in de goede richting gestuurd kan worden. Tijdens het toetsproces zelf is transparantie eveneens van belang. De kandidaat kan alleen optimaal presteren op een toets wanneer de scoringsregels duidelijk zijn en de kandidaat zijn antwoordstrategie daarop kan aanpassen. Bij meerkeuzevragen worden kandidaten (als het goed is) geïnstrueerd om altijd alle vragen te beantwoorden. Een onbeantwoorde vraag wordt immers als fout gerekend terwijl de kandidaat de kans heeft goed te gokken. Desondanks zijn er nog altijd kandidaten die vragen onbeantwoord laten. Als zo'n eenvoudige scoringsregel al zo moeilijk is uit te leggen, hoeveel lastiger wordt het dan niet met de ingewikkelde regels die nodig zijn voor een psychometrisch verantwoorde scoring van innovatieve vraagvormen! Laat staan wanneer voor elke vraag weer een nieuwe regel geldt en dus een nieuwe antwoordstrategie nodig is.

Geen tegenstander van digitaal toetsen

Verheul gaf aan bepaald geen tegenstander te zijn van digitaal toetsen. Integendeel, het is een vorm die een didactische meerwaarde kan hebben, maar alleen als vraagvormen worden gebruikt die op zichzelf een meerwaarde hebben, dat wil zeggen vraagvormen die niet op papier kunnen worden afgenomen. Hot-spotvragen zijn daar een goed voorbeeld van en vragen waarbij media functioneel zijn. Verheul stelde voor om naar de genoemde punten gericht onderzoek te doen zodat discussies hierover niet meer alleen op aannames worden gebaseerd.

De quotiëntregel voor de multiple responsevraag

Dat het gebruik van innovatieve vraagvormen nog in de kinderschoenen staat, bewijzen de scoringsregels die worden toegepast bij de multiple responsevraag. Ervaringen uit de praktijk laten zien dat het gebruik van een goede scoringsregel niet zo eenduidig is en niet zo eenvoudig ligt. Het artikel van Vos elders in dit nummer gaat daar nader op in. Voor de deelnemers aan de studiemiddag werd duidelijk dat het scoren van innovatieve vraagvormen complex is. Een scoringsregel bedenken in een toetssysteem is werk voor specialisten.

Innovatieve vraagvormen

Draaijer gaf een toelichting op de manier waarop binnen de Vrije Universiteit wordt omgegaan met digitale innovatieve vraagvormen.

De gokkans bij meerkeuzevragen is door de jaren heen vaak als bezwaar genoemd. Bovendien leent deze vraagvorm zich niet voor het toetsen van elk soort leerdoel. Innovatieve vraagvormen vormen een (mogelijk) antwoord op de problemen rond het gebruik van meerkeuzevragen. Het vernieuwend karakter van innovatieve vraagvormen bestaat uit:

- het format: het soort respons dat de kandidaat moet geven (bijvoorbeeld: kiezen of zelf formuleren);
- de wijze waarop de respons wordt gegeven (keyboard, muis);
- het gebruik van media;
- het niveau van interactiviteit: de mate waarin de respons van de kandidaat weer leidt tot een 'reactie' van de vraag (feedback bijvoorbeeld);
- de scoringswijze.

Most Constrained		Intermediate Constraint Item Types					Least Constrained
Less Complex	Fully Selected						Fully Constructed
	1. Multiple Choice	2. Selection/ Identification	3. Reordering/ Rearrangement	4. Substitution/ Correction	5. Completion	6. Construction	7. Presentation/ Portfolio
	1A. True/False (Haladyna, 1994c, p.54)	2A. Multiple True/False (Haladyna, 1994c, p.58)	3A. Matching (Osterlind, 1998, p.234; Haladyna, 1994c, p.50)	4A. Interlinear (Haladyna, 1994c, p.65)	5A. Single Numerical Constructed (Parshall et al, 2002, p. 87)	6A. Open-Ended Multiple Choice (Haladyna, 1994c, p.49)	7A. Project (Bennett, 1993, p.4)
	1B. Alternate Choice (Haladyna, 1994c, p.53)	2B. Yes/No with Explanation (McDonald, 2002, p.110)	3B. Categorizing (Bennett, 1993, p.44)	4B. Sore-Finger (Haladyna, 1994c, p.67)	5B. Short-Answer & Sentence Completion (Osterlind, 1998, p.237)	6B. Figural Constructed Response (Parshall et al, 2002, p.87)	7B. Demonstration, Experiment, Performance (Bennett, 1993, p.45)
	1C. Conventional or Standard Multiple Choice (Haladyna, 1994c, p.47)	2C. Multiple Answer (Parshall et al, 2002, p.2; Haladyna, 1994c, p.60)	3C. Ranking & Sequencing (Parshall et al, 2002, p.2)	4C. Limited Figural Drawing (Bennett, 1993, p.44)	5C. Cloze-Procedure (Osterlind, 1998, p.242)	6C. Concept Map (Shavelson, R. J., 2001; Chung & Baker, 1997)	7C. Discussion, Interview (Bennett, 1993, p.45)
More Complex	1D. Multiple Choice with New Media Distractors (Parshall et al, 2002, p.87)	2D. Complex Multiple Choice (Haladyna, 1994c, p.57)	3D. Assembling Proof (Bennett, 1993, p.44)	4D. Bug/Fault Correction (Bennett, 1993, p.44)	5D. Matrix Completion (Embretson, S, 2002, p. 225)	6D. Essay (Page et al, 1995, 561-565) & Automated Editing (Breland et al, 2001, pp.1-64)	7D. Diagnosis, Teaching (Bennett, 1993, p.4)

Figuur 1 Taxonomie voor vraagtypen (Scalise & Gifford, 2006)

Uit de **praktijk**

Voor de keuze en vervolgens de constructie van vraagtypen bevelen Draaijer en Hartog (2007) het gebruik van ontwerppatronen – *design patterns* – aan. Een dergelijk patroon geeft een overzicht van contexten (of combinaties van contexten) waarin bepaalde vraagtypen gebruikt kunnen worden, een taxonomische aanduiding van de vraagtypen, mogelijke formats of delen daarvan en uitgewerkte voorbeelden.

Een goed bruikbare taxonomie voor vraagtypen is die van Scalise & Gifford (2006). Deze ontwerppatronen worden op de VU gebruikt om docenten te helpen rationele keuzes te maken bij het samenstellen van hun toetsen en uit verschillende vraagtypen. Wat er nog aan de patronen ontbreekt zijn scoringsregels en psychometrische waarden. Draaijer demonstreerde voorbeelden van items die met behulp van *design patterns* geconstrueerd zijn.

Nieuwe vraagvormen in TestVision

Molkenboer en Soeting gaven aan dat nieuwe vraagvormen nieuwe toetsmogelijkheden bieden mits ze goed worden gebruikt. Molkenboer liet aan de hand van een aantal in de praktijk gebruikte innovatieve vraagvormen van verschillende organisaties zien wat daarmee mogelijk is. Deze vraagvormen kunnen kennis/inzicht en zelfs toepassingen meten hetgeen met traditionele vraagvormen niet mogelijk is. Soeting gebruikte voorbeelden waaruit bleek dat het niet altijd gewenst is om het aantal goede antwoorden te vermelden bij een multiple responsevraag. In de praktijk weet een arts bijvoorbeeld vooraf ook niet welke ziekten (hoeveel dus) uit te sluiten zijn op basis van zijn diagnose. In dat geval is het relevant om te toetsen of een kandidaat weet welke ziekten afvallen op basis van de diagnose en wordt het aantal goede antwoorden dus bewust niet genoemd. Het is aan de vraagontwikkelaar om te bepalen of in de stam het aantal goede antwoorden al dan niet wordt genoemd. Het aantal goede antwoorden is geen eigenschap van de multiple responsevraag maar een instrument in de handen van de vraagontwikkelaar. Tienduizenden kandidaten hebben inmiddels toetsen met verschillende vraagvormen gemaakt. Negatieve opmerking van kandidaten over het gebruik van verschillende vraagvormen zijn niet ontvangen. Het tegenovergestelde was wel het geval. Kandidaten vonden toetsen met nieuwe vraagvormen veel meer praktijkgericht, minder theoretisch en vooral leuker en prikkelender (ook al was de kandidaat gezakt). Ook schakelen tussen verschillende vraagvormen leverde voor kandidaten geen problemen op. Wel hadden ze eerst een oefentoets gemaakt zodat ze bekend waren met de wijze waarop nieuwe vraagvormen beantwoord moesten worden. Maar toegegeven, gericht onderzoek is er niet gedaan.

RESULTAAT VAN DE STUDIEMIDDAG

Aanbevelingen voor kandidaten:

- geef ze van te voren een oefentoets;
- geef ze informatie over hoe een vraag gescoord wordt;
- geef aan hoe ze een multiple responsevraag moeten beantwoorden:
 - 'Let op! Meer antwoorden kunnen goed zijn.'
 - 'Let op! Er zijn 2 goede antwoorden.'

Aanbevelingen voor vraagontwikkelaars:

- leer ze hoe ze goede vragen moeten construeren;
- geef docenten gelegenheid zich te bekwamen in het maken van de correcte vragen.

Aanbevelingen voor onderzoekers:

- onderzoek naar effecten van innovatieve vraagvormen is gewenst.

Literatuur

- Draaijer, S. & Hartog, R. (2007). Design Patterns for digital item types in Higher Education. *e-Journal of Instructional Science and Technology*, Accepted for publication.
- Scalise, K. & Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. *The Journal of Technology, Learning, and Assessment*, 4(6), 4-44, m.n. 9.

□ Ab van Beek was dagvoorzitter ad interim.