

Optimale **uniforme scoringsregels** voor innovatieve vraagvormen

□ Hans Vos, Marlies Kloppenburg en Onno Tomson

Gebruikers van toetssystemen rekenen erop dat deze betrouwbaar zijn en uitgaan van de optimale, uniforme scoringsregels voor innovatieve vraagvormen. De discussie over de quotiëntregel en de ontwikkeling van de productregel voor scoring van multiple responsvragen, waarbij het aantal goede antwoorden niet wordt weergegeven, heeft eerder in dit tijdschrift laten zien waar knelpunten liggen. Dit artikel gaat verder in op scoringsregels. Hans Vos houdt zich aanbevolen voor reacties.

Een terugblik

Lampe en Eggen (2005) hebben laten zien dat op item-niveau bij een evenwichtige verdeling over de verschillende kennisniveaus de dichotome scoring doorgaans de hoogste betrouwbaarheid geeft. Op toetsniveau geeft dichotome scoring bij gemengde toetsen geen optimale score. In november 2006 vond via de TOETS-HO-Lijst een e-mail-discussie plaats over het gebruik van innovatieve vraagvormen die leidde tot een studiemiddag over dit onderwerp. Hiervan is een verslag verschenen in EXAMENS (2008) en Hans Vos publiceerde een uitwerking van de quotiëntregel (Vos & de Graaf, 2008). In een reactie op het ontwerp van de quotiëntregel formuleerde Terwisscha van Scheltinga (2009) de productregel. Bovendien heeft Vos op verzoek van Tomson namens ECABO verder onderzoek uitgevoerd om ook voor andere innovatieve vraagvormen wetenschappelijk onderbouwde scoringsregels te ontwikkelen. Dit artikel geeft het onderzoeksresultaat op hoofdlijnen weer. Het onderzoeksrapport is op te vragen bij Vos.

Innovatieve vraagvormen

Onder innovatieve vraagvormen worden gesloten vraagvormen verstaan zoals de multiple responsvraag, ordeningsvraag, matchvraag, matrixvraag en de multiple true/false vraag. Het gebruik van innovatieve vraagvormen komt door de opkomst van elektronische toetssystemen steeds vaker voor. Elektronische toetssystemen scoren deze vraagvormen automatisch. Het komt voor dat een toets met verschillende innovatieve vraagvormen gemaakt met toetssysteem A een andere totaalscore oplevert dan gemaakt met toetssysteem B. Een situatie die niet gewenst is.

Scoring

Vragen kunnen zowel dichotoom als polytoom gescoord worden. Bij dichotoom scoren krijgt het antwoord een score '1' als het item volledig juist is beantwoord en anders een score '0'. Bij een polytome scoring zijn ook tussenliggende waarden mogelijk, afhankelijk van het gegeven antwoord. True/false vragen en de kort antwoord- of invulvraag zijn per definitie dichotoom. Deze vraagvormen kunnen alleen maar óf goed óf fout beantwoord worden. Innovatieve vraagvormen kunnen dichotoom of polytoom worden gescoord. De hier behandelde scoringsregels hebben betrekking op de polytome scoring.

Bij polytome vraagvormen kan een antwoord gedeeltelijk goed zijn, maar ook gedeeltelijk fout. Als een kandidaat de kleuren van de Nederlandse vlag moet selecteren uit de vijf kleuren rood, wit, blauw, oranje en geel en hij geeft als antwoord rood en wit, dan is dat antwoord gedeeltelijk goed. Kiest hij wit, blauw en oranje dan is het antwoord gedeeltelijk goed én gedeeltelijk fout. Het is reëel om voor het foute antwoord aftrek van punten te geven. In feite wordt hiermee dus de traditionele gokkans (bij dichotome vragen) in de scoringsregel van polytome vragen al min of meer verwerkt.

Multiple responsvraag

De multiple responsvraag is geïndiceerd voor het meten van (partiële) kennis over een aantal bij elkaar behorende



Bij polytome vraagvormen kan een antwoord gedeeltelijk goed zijn, maar ook gedeeltelijk fout

onderwerpen (bijvoorbeeld de kleuren van de Nederlandse vlag selecteren uit vijf kleuren). In de stam bij de vraag kan al dan niet worden aangegeven hoeveel juiste alternatieven moeten worden gekozen.

Het aantal juiste alternatieven wordt in de stam aangegeven

Vos & de Graaf (2008) stellen voor dit type vraag de quotiëntregel voor. De quotiëntregel luidt:

$$S_{\text{quotiëntregel}} = \frac{\text{proportie goed}}{1 + \text{proportie fout}}$$

Het aantal juiste alternatieven wordt niet in de stam aangegeven

Als het aantal juiste alternatieven niet wordt aangegeven, kan de kandidaat 'blind' alle alternatieven selecteren en kent de quotiëntregel altijd 0.5 punt toe. Immers, in deze situatie geldt: proportie goed = proportie fout = 1. Terwisscha van Scheltinga (2009) heeft de volgende productregel voorgesteld om aan dit bezwaar tegemoet te komen.

$$S_{\text{productregel}} = \text{proportie goed} * \left\{ 1 - \text{proportie fout} * \left[\frac{1}{a} + \left(\frac{n-k}{n} \right) \left(1 - \frac{1}{a} \right) \right] \right\}$$

waarbij:

n = het totale aantal alternatieven;

k = het aantal juiste alternatieven;

a = weegfactor voor het meetellen van foute antwoorden ($a \geq 1$) (experimenteel is vastgesteld dat $a=2$ het beste resultaat oplevert).

Aanbeveling

De betrouwbaarheden en de discriminerende vermogens van de productregel en de quotiëntregel leveren vergelijkbare waarden op. Als het aantal juiste alternatieven is aangegeven in de stam wordt de quotiëntregel geprefereerd boven de productregel vanwege transparantie-overwegingen.

Orderingsvraag

Bij orderingsvragen moet een aantal 'objecten' in een juiste volgorde worden geplaatst (bijvoorbeeld een aantal Amerikaanse presidenten in een chronologische volgorde). Eggen & Lampe (2005) hebben vier verschillende scoringsmethoden voor orderingsvragen met K vrij te ordenen objecten (minstens 3 objecten) met elkaar vergeleken.

Hieronder wordt een scoringsvoorbeeld gegeven met vier te ordenen objecten met als juiste volgorde ABCD. Stel dat de kandidaat als antwoord 'ABDC' geeft. Hoe kan dit worden gescoord?

1. Tel de juiste objecten op absolute positie (maximaal K punten te behalen, dus hier 4 punten). Het antwoord 'ABDC' levert 2 punten op, omdat de objecten 'A' en 'B' op de juiste positie staan.
2. Tel de juiste ordening van direct naaste paren (maximaal K-1 punt te behalen, dus hier 3 punten). Het antwoord 'ABDC' levert 1 punt op, omdat de onderlinge volgorde van 'AB' juist is en die van 'BD' en 'DC' onjuist zijn.
3. Tel de juiste ordening van naaste paren (maximaal K-1 punt te behalen, dus hier 3 punten). Het antwoord 'ABDC' levert 2 punten op, omdat de onderlinge volgorde van 'AB' en 'BD' juist zijn en die van 'DC' onjuist is.
4. Tel alle juiste ordeningen (maximaal $K*(K-1)/2$ punt te behalen, dus hier 6 punten).
5. Het antwoord 'ABDC' levert 5 punten op, omdat de onderlinge volgorde van 'AB', 'AD', 'AC', 'BD' en 'BC' juist zijn en alleen de onderlinge volgorde van 'DC' onjuist is.

De door Eggen & Lampe vergeleken scoringsmethoden zijn doorgerekend met de quotiëntregel. Hierbij gaf de bovengenoemde vierde scoringsmethode de hoogste betrouwbaarheden en discriminerende vermogens.

De quotiëntregel voor de vierde methode luidt:

$$S_{\text{quotiëntregel alle juiste ordeningen}} = \frac{\text{proportie goed}}{1 + \text{proportie goed bij raden}} = \frac{\frac{S}{K(K-1)/2}}{1 + \frac{\frac{S}{K(K-1)/2}}{K(K-1)/2}} = \frac{S}{K(K-1) - S}$$

Waarbij:

K = vrij te ordenen objecten;

S = aantal juiste volgorde van alle mogelijke paren van objecten.

Aanbeveling

Hanteer als scoringsmethode alle juiste ordeningen in combinatie met de quotiëntregel.

Als het aantal juiste alternatieven is aangegeven in de stam wordt de quotiëntregel geprefereerd



Matchvraag

Bij een gelijke en ongelijke matchvraag moeten steeds twee zaken met elkaar worden gecombineerd. Dit vraagtype is geïndiceerd bij het toetsen van verbanden leggen (bijvoorbeeld vier steden koppelen aan vier provincies). Lampe & Eggen (2006) scoren matchvragen door voor iedere juist gelegde relatie 1 punt toe te kennen (dus zonder aftrek voor verkeerd gelegde relaties). Teelen Kennismanagement (2006) houdt in haar scoringsregel wel rekening met aftrek voor verkeerd gelegde relaties. De quotiëntregel blijkt na onderzoek ook bij de matchvraag de hoogste betrouwbaarheid en het hoogste discriminerend vermogen te hebben.

$S_{\text{quotiëntregel ongelijke matching}} =$

$$\frac{\text{proportie goed}}{1 + \text{proportie goed bij raden}} = \frac{\frac{S}{K}}{1 + \frac{K-S}{K(N-1)}} = \frac{\frac{S}{K}}{1 + \left(1 - \frac{S}{K}\right)\left(\frac{1}{N-1}\right)}$$

Waarbij:

K = aantal startalternatieven;

N = aantal koppelalternatieven;

S = aantal goed gelegde koppelingen.

Aanbeveling

Hanteer als scoringsmethode de quotiëntregel bij ongelijke matching omdat die ook voor gelijke matching toepasbaar is.

Matrixvraag

Bij de matrixvraag moet van een aantal gegevens worden aangegeven welke kenmerken van toepassing zijn. Dit type vraag is geïndiceerd als men kennis van diverse aspecten van een onderwerp in één vraag wil toetsen. De matrixvraag is ook op te vatten als een variant op de matchvraag, waarbij startalternatieven (gegevens) moeten worden gekoppeld aan alle juiste koppelalternatieven (kenmerken). Een voorbeeld van een matrixvraag is dat van de steden Groningen, Arnhem en Zutphen moet worden aangegeven of zij een provinciehoofdstad zijn, meer dan 100.000 inwoners hebben en aan een grote rivier liggen. De matrixvraag kan beschouwd worden als een verzameling onafhankelijke multiple responsvragen (Lampe & Eggen, 2006; Teelen, 2006). Eerder is aangegeven dat de productregel de optimale scoringsregel is voor iedere

afzonderlijke multiple responsvraag als het aantal juiste koppelalternatieven niet is aangegeven voor ieder startalternatief en de kandidaat dus ieder startalternatief 'blind' aan alle koppelalternatieven kan koppelen (overal 'ja' aankruisen in de matrix). De toegekende score is dan de gemiddelde score over alle afzonderlijke multiple responsvragen.

Stel dat de kandidaat in het voorbeeld voor Groningen, Arnhem en Zutphen de volgende scores realiseert voor ieder van de drie afzonderlijke multiple responsvragen (volgens de productregel (met weegfactor $a = 2$)):

Groningen: $1/6$;

Arnhem: $2/3$;

Zutphen: $7/12$.

Als score voor de totale matrixvraag wordt dan de gemiddelde productscore toegekend:

$$(1/6 + 2/3 + 7/12)/3 = 17/36.$$

Aanbeveling

Hanteer de gemiddelde score op basis van de productregel over de afzonderlijke multiple responsvragen.

Multiple true/false

Bij een multiple true/false vraag moet van meerdere beweringen (proposities) bij een stam worden aangegeven of deze juist of onjuist zijn. Dit type vraag is geïndiceerd bij onder meer het herkennen van begrippen. Er is een ondubbelzinnig juist antwoord en er zijn weinig plausibele alternatieven. De multiple true/false vraag is op te vatten als een speciaal geval van de (gelijke of ongelijke) matchvraag, waarbij er K startalternatieven zijn in de vorm van K true/false beweringen en er $N=2$ koppelalternatieven (true of false) zijn.

$$S_{\text{MTF}} = S_{\text{quotiëntregel (ongelijke) matching met } N=2} = \frac{\frac{S}{K}}{1 + \left(1 - \frac{S}{K}\right)\left(\frac{1}{N-1}\right)} = \frac{\frac{S}{K}}{1 + \left(1 - \frac{S}{K}\right)} = \frac{\text{proportie goed}}{1 + \text{proportie fout}}$$

Waarbij:

K = totaal aantal beweringen;

S = aantal goed aangekruiste beweringen.



De quotiëntregel heeft ook bij de matchvraag de hoogste betrouwbaarheid en discriminerend vermogen

Aanbeveling

Hanteer als scoringsmethode de quotiëntregel bij ongelijke matching met $N=2$.

Verzoek aan de lezers

Het verzoek aan de lezers is om de scoringsregels kritisch te bestuderen en een reactie te sturen naar de auteurs. Het doel is om te komen tot optimale en uniform te gebruiken scoringsregels voor innovatieve gesloten vraagvormen. Ontwikkelaars van de volgende in Nederland veel gebruikte elektronische toetsystemen worden met name uitgenodigd te reageren: Egel, Etude, Question Mark Perception, TestVision, Wintoets. Maar ook reacties van ontwikkelaars van Learning Managementsystemen met toetsfunctionaliteiten zijn welkom. Het resultaat kan zijn dat uiteindelijk de optimale scoringsregels in alle systemen worden opgenomen.

Literatuur

- Eggen, T. & Lampe, T. (2005). Een kwestie van volgorde: de scoring van ordeningsvragen. *Paper ORD 2005*. Gent, België.
- Lampe, T. & Eggen, T. (2005). Hoe scoor je de antwoor-

den? Meervoudige keuzevragen. *EXAMENS*, 2(2), 18-21.

Lampe, T. & Eggen, T. (2006). *Scoring van matchingvragen: Onderzoek naar de betrouwbaarheid van verschillende scoringsmethoden*. Arnhem: Cito.

Teelen Kennismanagement (26 oktober 2006). *Scoring matchingvraag en rangschikvraag*. Notitie uitgereikt op studiemiddag (14-07-07). Gebruikersverenigingen QMP en TestVision.

Terwisscha van Scheltinga, E. (2009). Productregel als scoringsmethode voor meervoudige keuzevragen. *EXAMENS*, 6(1), 26-27.

Vos, H.J. & de Graaf, A.M. (2008). Een nieuwe scoringsmethode voor meervoudige keuzevragen. *EXAMENS*, 5(1), 18-21.

□ De heer dr. H.J. Vos is werkzaam aan de Universiteit Twente, Faculteit Gedragwetenschappen, afdeling Onderzoeksmethodologie, Meetmethoden en Data-analyse. Mw drs. E.F.M. Kloppenburg en de heer O. Tomson zijn werkzaam bij ECABO als Adviseur examinering respectievelijk Beheerder informatiesystemen afdeling Examinering. E-mail: h.j.vos@gw.utwente.nl.

Ook de examens van



staan onder
toezicht van de

Examen
kamer
kwaliteit door onafhankelijkheid