

De quotiëntregel

□ Hans Vos en Anne-Marie de Graaf

De quotiëntregel is op de studiedag over innovatieve vraagvormen gepresenteerd als een nieuwe methode voor het optimaal scoren van meervoudige keuzevragen. Dit type vraag is een variant op de klassieke meerkeuzevraag, waarbij meer goede antwoorden mogelijk zijn. De quotiëntregel komt voort uit een vergelijkend onderzoek naar twee scoringsregels in opdracht van het Nederlands bureau brandweerexamens (Nbbe). Het Nbbe heeft de intentie de quotiëntregel in gebruik te nemen.

Nieuwe toetsen

Toen het Nederlands bureau brandweerexamens in april 2006 van papier-en-pen-toetsen (met alleen klassieke meerkeuzevragen) overging op online-toetsing met behulp van het pakket TestVision, bleken de slagingspercentages sterk te zijn gedaald voor toetsen waarin relatief veel meervoudige keuzevragen waren opgenomen. De meervoudige keuzevraag (multiple response item) is een variant op de klassieke meerkeuzevraag en wordt vaak gebruikt wanneer er meer juiste antwoordmogelijkheden zijn. Deze gesloten vraagvorm is daarom met name geïndiceerd voor het meten van de kennis van kandidaten over een aantal bij elkaar horende onderwerpen.

Voorbeeld 1 Meervoudige keuzevraag zonder vermelding van aantal goede antwoorden

Welke kleuren heeft de Nederlandse vlag? (*juist)

- [A*] blauw
- [B] geel
- [C] oranje
- [D*] rood
- [E*] wit

Het kunnen meten van verschillende kennisniveaus (partiële kennis) is een belangrijk voordeel van de meervoudige keuzevraag ten opzichte van de klassieke meerkeuzevraag. In voorbeeld 1 is het aantal goede antwoorden niet aangegeven. Maar soms wordt dit wel aangegeven in de stam, zoals in voorbeeld 2, dat is ontleend aan de oefentoets voor een aankomend brandweerman of vrouw op de Nbbe-site (www.nbbe.nl).

Voorbeeld 2 Meervoudige keuzevraag met vermelding van aantal goede antwoorden

Een beginnende backdraft kunt u herkennen aan een aantal verschijnselen. Welke drie van onderstaande signalen wijzen op een beginnende backdraft?

- [A] stilte
- [B*] fluitende geluiden
- [C] helder rode vlammen
- [D*] kolkende rook
- [E] blauwe vlammen in de rooklaag
- [F*] aanwezigheid smeulende schuimmatrassen

Het Nbbe zag – naast andere factoren – de relatieve strenge scoringsregel die men hanteerde als mogelijke verklaring voor deze sterke daling van de slagingspercentages. Het Nbbe gebruikte als scoringsregel: proportie goed minus proportie fout met als ondergrens 0. Men vroeg zich af of de regel 'proportie goed' beter zou zijn. Om een antwoord op deze vraag te krijgen benaderde het Nbbe de eerste auteur van dit artikel en vroeg hem de twee scoringsregels met elkaar te vergelijken.

Een meervoudige keuzevraag is een variant op de klassieke meerkeuzevraag



Scoringsmethoden op itemniveau

De fundamentele vraag is hoe de meervoudige keuzevraag het best gescoord kan worden, omdat er vaak verschillende scoringsmogelijkheden zijn geïmplementeerd in pakketten voor computergestuurd toetsen (o.a. TestVision en QMP). Al eerder zijn in EXAMENS de volgende vier scoringsmethoden besproken (Lampe & Eggen, 2005):

1. *Dichotoom*: een kandidaat krijgt alleen score '1' als hij alle juiste alternatieven selecteert. In alle andere gevallen krijgt hij score '0';
2. *Polytoom*: voor elk geselecteerd juist alternatief krijgt een kandidaat een punt;
3. *Polytoom met aftrek*: voor elk geselecteerd juist alternatief krijgt een kandidaat een punt, maar voor elk geselecteerd fout alternatief wordt een punt afgetrokken ('straf');
4. *Polytoom met aftrek en ondergrens*: zelfde als methode 3 hiervoor, alleen kan een kandidaat niet een (vaak ongewenste) negatieve score krijgen. Er wordt een ondergrens gehanteerd van bijvoorbeeld '0' punten.

Lampe & Eggen (2005) laten zien dat de dichtome scoringsmethode doorgaans de hoogste betrouwbaarheid heeft bij een evenwichtige verdeling over de verschillende kennisniveaus. Deze methode is echter niet geschikt voor het toekennen van deelscores aan personen met verschillende kennisniveaus. Daarvoor is de 'normale' polytome scoringsmethode de meest efficiënte en transparante scoringsmethode met de hoogste betrouwbaarheid. De tweede en de derde scoringsmethode blijken equivalent te zijn, omdat ze feitelijk lineaire transformaties van elkaar zijn. De laatste polytome scoringsmethode (4) ten slotte wordt door Lampe & Eggen (2005) als nodeloos gecompliceerd omschreven en qua betrouwbaarheid altijd wat lager dan de beste scoringsmethode.

Scoringsmethoden op toetsniveau

De vier verschillende scoringsmethoden waren door Lampe & Eggen (2005) primair bedoeld voor gebruik op itemniveau en dus uitdrukkelijk niet voor gebruik op toetsniveau. Als meervoudige keuzevragen echter worden opgenomen in een gemengde toets (bijvoorbeeld een toets bestaande uit klassieke meerkeuzevragen én meervoudige keuzevragen), heeft

dit als consequentie dat de items moeten worden gewogen om een totaalscore te kunnen toekennen. In het toetspakket waarmee het Nbbe werkt voor haar toetsen, is hiermee op een praktische manier omgegaan. Bij de drie polytome methoden worden geen punten toegekend of afgetrokken voor ieder goed of fout geselecteerd alternatief, maar er wordt gerekend met proporties van goede en foute alternatieven. De scoretoekenning voor de drie polytome methoden wordt dan dus 'proportie goed' voor methode (2), 'proportie goed minus proportie fout' voor methode (3) en 'proportie goed minus proportie fout met ondergrens 0' voor methode (4) als dit verschil negatief wordt. Wiskundig gezien zijn de scores op deze manier voor methoden (2) en (4) genormeerd in het standaardinterval $[0,1]$ en voor methode (3) in het interval $[-1,1]$. Korthedshalve zal de 'gewone' genormeerde polytome regel (2) in de rest van dit artikel worden aangeduid als de 'C(ito)-regel' en de genormeerde polytome regel met aftrek en ondergrens (4) als de 'T(estVision)-regel'.

Aantal goede alternatieven

In de vier door Lampe & Eggen (2005) besproken methoden wordt ervan uitgegaan dat de kandidaat in de stam wordt verteld hoeveel goede alternatieven er zijn. Omdat in sommige (praktijk)situaties de kandidaat in staat moet zijn om uit de aangeboden keuzes zelf te bepalen hoeveel alternatieven goed zijn, wil het pakket TestVision echter niet verplicht laten vastleggen hoeveel antwoorden goed zijn. Bij keuze voor een genormeerde polytome regel, komt de C-regel echter niet in aanmerking als het aantal goede alternatieven niet wordt verteld in de stam. Immers, door alle alternatieven aan te kruisen is de kandidaat er bij scoring met deze regel altijd van verzekerd dat alle goede alternatieven hier ook bij zitten en dus krijgt hij automatisch de maximale score van 1 toegekend. Bovendien zou door veel alternatieven te selecteren de kansscore erg hoog worden bij gebruik van de C-regel als het aantal goede alternatieven niet wordt vermeld. Als de kandidaat alle alternatieven selecteert (en deze niet allemaal goed zijn), is zowel zijn 'proportie goed' als zijn 'proportie fout' gelijk aan 1 en krijgt hij dus altijd de minimale score van 0 toegekend bij gebruik van de T-regel. Deze score wordt overigens



Deze gesloten vraagvorm is geïndiceerd voor het meten van kennis van kandidaten over een aantal bij elkaar horende onderwerpen

ook toegekend als bijvoorbeeld alle alternatieven op één na goed zijn, waarvan men zich kan afvragen of dit altijd wenselijk is. Als alle alternatieven goed zijn, is zijn 'proportie goed' gelijk aan 1 en zijn 'proportie fout' gelijk aan 0. Evenals bij gebruik van de C-regel, wordt bij gebruik van de T-regel in dit geval dus de maximale score van 1 toegekend.

In de volgende paragraaf wordt als compromis voor de C-regel en de T-regel de quotiëntregel geformuleerd, nader uitgewerkt en toegelicht aan de hand van een voorbeeld.

Quotiëntregel als compromis

De quotiëntregel kent als score toe: $(\text{proportie goed}) / (1 + \text{proportie fout})$. Deze scoringsregel ligt qua scoretoekenning tussen de C-regel en de T-regel in. Het is een compromisregel die een aantal dezelfde eigenschappen heeft als deze beide regels, namelijk:

1. genormeerd in het standaardinterval $[0,1]$;
2. minimale score van 0 toekent als er geen enkel goed alternatief wordt geselecteerd;
3. 'proportie goed' toekent als alleen goede alternatieven worden geselecteerd;
4. maximale score van 1 toekent als alleen alle goede alternatieven worden geselecteerd (feitelijk een speciaal geval van gewenste eigenschap (3)).

Bovendien is een variant op deze compromisregel toepasbaar als in de stam het aantal goede alternatieven niet is aangegeven en de kandidaat dus in principe alle alternatieven kan selecteren. Aangevoerd kan worden dat deze regel nooit strenger is dan de T-regel en altijd minstens even streng als de C-regel. De scoretoekenning bij de quotiëntregel is als volgt:

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

Ter illustratie van deze regel: Stel dat de kandidaat in voorbeeld 1 over de kleuren van de Nederlandse vlag (met 'totaal aantal alternatieven' = 5 en 'totaal aantal goede alternatieven' = 3) in totaal 4 alternatieven selecteert en wel de alternatieven A*BCD*, dan heeft de kandidaat dus 2 goede alternatieven geselecteerd en 2 foute alternatieven ('afleiders') geselecteerd. De 'proportie goed' is dus gelijk aan: $(\text{aantal goed}) / (\text{totaal}$

aantal goed) = $2/3$. En de 'proportie fout' is gelijk aan: $(\text{aantal fout}) / (\text{totaal aantal fout}) = 2 / (5 - 3) = 1$.

De quotiëntregel zou in dit geval een score toekennen van:

$$s_{\text{quotiëntregel}} = \frac{\text{proportie goed}}{1 + \text{proportie fout}} = (2/3) / (1 + 1) = 1/3$$

De C-regel zou in dit geval een score toekennen van:

$$s_{\text{C-regel}} = \text{'proportie goed'} = 2/3.$$

Ten slotte zou de T-regel in dit geval een score toekennen van:

$$s_{\text{T-regel}} = (\text{proportie goed} - \text{proportie fout}) = (2/3) - 1 = -1/3$$

en vanwege de ondergrens '0' zou de T-regel in dit geval dus 0 punt toekennen.

In dit voorbeeld geldt dus dat de quotiëntregel een compromis is tussen de C-regel en de T-regel. Verder is eenvoudig na te gaan dat de quotiëntregel de vier eerder geformuleerde eigenschappen deelt met de C-regel en de T-regel. Merk hierbij op dat de quotiëntregel (evenals de C-regel) op een 'natuurlijke' manier is genormeerd in het standaardinterval $[0,1]$. In tegenstelling tot bijvoorbeeld de T-regel, hoeft er dus niet op een min of meer arbitraire manier een ondergrens te worden gesteld.

Uiteraard moeten we ook nog in zijn algemeenheid bewijzen dat de quotiëntregel een compromis is tussen de C-regel en de T-regel:

$$0 \leq s_{\text{T-regel}} \leq s_{\text{quotiëntregel}} \leq s_{\text{C-regel}} \leq 1.$$

Eenvoudig is na te gaan dat de drie regels dezelfde score toekennen in de volgende situaties:

$$1 \quad s_{\text{T-regel}} = s_{\text{quotiëntregel}} = s_{\text{C-regel}} = 0,$$

als geen enkel goed alternatief wordt gekozen (i.e., 'proportie goed' = 0);

$$2 \quad s_{\text{T-regel}} = s_{\text{quotiëntregel}} = s_{\text{C-regel}} = \text{'proportie goed'}$$

Het kunnen meten van verschillende kennisniveaus is een belangrijk voordeel van de meervoudige keuzevraag



als alleen goede alternatieven worden gekozen (i.e., 'proportie fout' = 0);

$$3 \quad s_{\text{T-regel}} = s_{\text{quotiëntregel}} = s_{\text{C-regel}} = 1,$$

als alleen alle goede alternatieven worden gekozen (i.e., 'proportie goed' = 1 en 'proportie fout' = 0).

Omdat bij de quotiëntregel de 'proportie goed' wordt gedeeld door een getal dat altijd groter is dan of gelijk is aan 1, volgt hieruit onmiddellijk dat de C-regel altijd een minstens even hoge score toekent als de quotiëntregel. Omdat verder bij de T-regel de 'proportie fout' wordt afgetrokken van de 'proportie goed' met als ondergrens 0, volgt hieruit onmiddellijk dat de C-regel ook altijd een minstens even hoge score toekent als de T-regel. We moeten dus alleen nog in z'n algemeenheid bewijzen dat de quotiëntregel altijd een minstens even hoge score toekent als de T-regel. Dit bewijs is iets ingewikkelder maar is voor geïnteresseerden opvraagbaar bij de eerste auteur van dit artikel.

Ten slotte bleek nog dat de quotiëntregel het gemiddeld bijna overal beter deed dan de T-regel en min of meer vergelijkbaar was met de C-regel wat betreft de beide door Lampe & Eggen (2003, 2005) voorgestelde psychometrische kwaliteitsindicatoren betrouwbaarheid en discriminerend vermogen.

Scoren bij selectie van alle alternatieven

Stel dat de kandidaat alle alternatieven selecteert (en deze niet allemaal goed zijn), dan heeft hij automatisch zowel alle goede als alle foute alternatieven geselecteerd en geldt uiteraard: 'proportie goed' = 'proportie fout' = 1. De quotiëntregel kent dan een score toe van:

$$s_{\text{quotiëntregel}} = \frac{\text{proportie goed}}{1 + \text{proportie fout}} = \frac{1}{1+1} = 0.5.$$

Hoewel niet zo hoog als bij de C-regel, die altijd een score van 1 oplevert door alle alternatieven te selecteren, lijkt deze gegarandeerde minimale score van 0.5 aan de (te) hoge kant te zijn en is dus niet realistisch. Als de kandidaat alle alternatieven selecteert, wordt

daarom voorgesteld om als score toe te kennen:

$$s_{\text{alle alternatieven}} = \frac{\text{totaal aantal goede alternatieven}}{\text{totaal aantal alternatieven}}$$

Eenvoudig is na te gaan dat bovenstaande regel is genormeerd in het standaardinterval [0,1] en een compromis is tussen de C-regel en de T-regel. Ook kent deze regel realistische scores toe:

1. Score neemt toe bij stijgend 'totaal aantal goede alternatieven' (met de maximale score '1' als alle alternatieven goed zijn) en neemt af bij dalend 'totaal aantal goede alternatieven' (met de theoretisch minimale score '0' als geen enkel alternatief goed is).
2. Als er in totaal evenveel goede als foute alternatieven zijn, dan wordt een score van 0.5 toegekend.
3. Als het 'totaal aantal goede alternatieven' meer/minder is dan de helft van het 'totaal aantal alternatieven', dan wordt een score groter/kleiner dan 0.5 toegekend.

In voorbeeld 1 zou deze regel dus een score toekennen van $3/5 = 0.6$. De toegekende score '(totaal aantal goede alternatieven)/(aantal goede alternatieven)' voor de situatie dat de kandidaat alle alternatieven selecteert, is in feite een toepassing van een variant op de quotiëntregel in de situatie die resulteert als proporties worden vervangen door aantallen [i.e., '(aantal goed)/(totaal aantal goed + aantal fout)'].

Literatuur

- Lampe, T. & Eggen, Th. (2003). Innovatieve itemtypes bij computertoetsing: het scoren van multiple response items. Paper op het 4e Nationale Examencongres van de NVE.
- Lampe, T. & Eggen, Th. (2005). Hoe scoor je de antwoorden? Meervoudige keuzevragen. *EXAMENS*, 2 (2), 18-21.

☐ Dr. ir. H.J. Vos is werkzaam als universitair docent aan de Universiteit Twente, Faculteit Gedragwetenschappen, afdeling Onderzoeksmethodologie, Meetmethoden en Data-analyse, Enschede. Mw. drs. J.M.P. de Graaf is werkzaam als beleidsmedewerker/ toetskundige bij het Nbbe (Nederlands bureau brandweereexamens).