

Klopt dit wel?



Een meerkeuzevraag moet vier alternatieven hebben

■ Silvester Draaijer

Dr. S. Draaijer is Onderwijsadviseur en Adviseur ICT en Onderwijs bij de Vrije Universiteit te Amsterdam.

■ Jackelien ter Burg

Drs. J. ter Burg is Toegepast Onderwijskundige en werkzaam als freelance toetsdeskundige.

Een meerkeuzevraag bestaat altijd uit een stam en een aantal alternatieven. Het aantal alternatieven kan variëren. In Nederland zijn vier alternatieven gebruikelijk, dus één correct antwoord en drie niet-correcte antwoorden. Maar is dat in alle gevallen wel de beste optie?

Richtlijnen voor aantal alternatieven

In hun review naar richtlijnen voor vraagconstructie vinden Haladyna, Downing en Rodriguez (2002) dat de richtlijnen over het aantal alternatieven verschillen. De meeste auteurs van handboeken over toetsconstructie adviseren om zo veel alternatieven als mogelijk op te nemen, andere auteurs hameren op vier alternatieven terwijl andere auteurs het gewenste aantal alternatieven simpelweg niet noemen. Een veelgehoorde reden om meer alternatieven op te nemen, is het verlagen van de raadkans op de vraag. Naarmate het aantal alternatieven groter is, wordt de raadkans immers kleiner.

Daardoor zal de hoeveelheid 'ruis' in de meting afnemen en zal men voor een gelijkblijvende betrouwbaarheid van de toets met minder vragen toekunnen. Omgekeerd wordt daarom vaak het advies gegeven om méér vragen in de toets op te nemen als er voor minder alternatieven wordt gekozen. Zie onderstaande tabel voor een richtlijn tussen het aantal alternatieven en de geadviseerde toetslengte.

Een nadeel van meer alternatieven is dat de leestijd voor de kandidaat toeneemt. Zie tabel 1 voor de relatie tussen het aantal alternatieven

Tabel 1. Richtlijnen per aantal alternatieven (Berkel en Bax 2002)

Aantal alternatieven	Minimaal aantal vragen in de toets	Noodzakelijke tijd
Twee	80	50 seconden
Drie	60	60 seconden
Vier of vijf	40	75 seconden

ven en de geadviseerde tijd. Hierbij maakt het natuurlijk uit of de alternatieven uit een getal, een woord of enkele zinnen bestaan. In het algemeen geldt dat hoe meer alternatieven er zijn, hoe minder snel de kandidaat het verschil tussen de alternatieven zal zien. Dat gaat dan mogelijk ten koste van de betrouwbaarheid van de toetsing omdat de kandidaat veel informatie in het werkgeheugen moet houden.

Aantrekkelijkheid niet-correcte antwoorden

Het advies om meer alternatieven in een toetsvraag op te nemen, is alleen nuttig onder de voorwaarde dat de niet-correcte alternatieven voldoende 'aantrekkelijk' zijn voor de minder goede kandidaat. Aantrekkelijke niet-correcte alternatieven zijn bijvoorbeeld misconcepties (Crouch & Mazur, 2001; Moelands, Noijons, & Rem, 1992). Alternatieven die de minder goede kandidaat zonder kennis van de stof direct kan wegstrepen, voegen niets toe aan de meetinformatie die de toets oplevert. Bij het vergroten van het aantal alternatieven neemt in dat geval de theoretische raatkans wel af, maar niet de 'empirische' raatkans. Dit wordt bevestigd door onderzoek waaruit is gebleken dat bij veel vierkeuzevragen slechts één of twee niet-correcte alternatieven effectief zijn (Paes & Cate, 2009; Rodriguez, 2005) en de andere helemaal niet. Uit onderzoek blijkt ook dat het in veel gevallen veel moeite kost om meer dan twee aannemelijke niet-correcte alternatieven te verzinnen (Mayenga, 2009). In lijn met onder andere Rodriguez (2005) lijkt de conclusie gerechtvaardigd te zijn dat een meerkeuzevraag met drie alternatieven - dus één correct antwoord en twee niet-correcte antwoorden - de optimale oplossing is.

Redenen voor minder of meer dan drie alternatieven

Bovenstaande conclusie hoeft niet te betekenen dat een meerkeuzevraag altijd drie alternatieven moet hebben. Een beter advies is om goed te onderzoeken wat binnen de eigen toetssituatie het optimale aantal is, zoals ook Haladyna et al.

(2002) suggereren. Er zijn mogelijk onderwerpen waarbij een tweede niet-correct antwoord dat wel aannemelijk is al heel moeilijk te verzinnen is. In dat geval past een tweekeuzevraag het beste, en dan bij voorkeur met een duidelijke vraagstelling met twee alternatieven die beide een antwoord op de vraag zijn ('alternate choice') en niet als juist/onjuist-vraag (Downing, 1992). Voor andere onderwerpen kan het gemakkelijk zijn om meer dan drie aannemelijke niet-correcte alternatieven te verzinnen.

In sommige gevallen is een vierde alternatief nodig om te voorkomen dat een kandidaat door te puzzelen tot het juiste antwoord kan komen. De figuur op pagina 54 betreft twee voorbeelden.

Het antwoord dat een kandidaat op de vraag wil geven, dient bij voorkeur bij de alternatieven te staan. Als een heel aannemelijk antwoord niet bij de alternatieven staat, is er een kans dat een kandidaat die het correcte antwoord eigenlijk niet weet naar het correcte antwoord wordt geholpen. Als je deze lijn volgt, is het te verantwoorden dat bijvoorbeeld binnen een examen voor een autoverkoper bij een vraag als 'Welke model voldoet het beste aan de wensen van deze klant' alle mogelijke automodellen als alternatieven zijn opgenomen.

Andere overwegingen

Ook bij een goed geconstrueerde meerkeuzevraag kan er een alternatief zijn dat heel weinig of niet gekozen wordt. Een dergelijk alternatief hoeft niet per definitie verwijderd te worden. Door een alternatief niet te kiezen, laat een kandidaat ook zien dat hij partiële kennis heeft. De vraagconstructeur kan hierdoor met de moeilijkheidsgraad van de vraag spelen. Immers, door meer of minder aannemelijke niet-correcte antwoorden op te nemen (zowel wat betreft aantal als aantrekkelijkheid) wordt de gemeten moeilijkheidsgraad, gemeten als percentage correct beantwoord, beïnvloed.

Uit de kwaliteitseisen bruikbaarheid of transparantie (vanuit de beleidsregels bijvoorbeeld) kan een besluit volgen om het aantal alternatieven

in een toets of vragenbank constant te houden. Bijvoorbeeld in het geval van een papieren toets met een apart antwoordformulier kan binnen bepaalde doelgroepen een wisselend aantal alternatieven ongewenste fouten veroorzaken. Het laatste antwoord kan immers antwoord C zijn, maar ook antwoord E of misschien wel een keer antwoord F.

In sommige digitale toetssystemen moet het aantal alternatieven in een toets constant zijn, bijvoorbeeld omdat voor een bepaalde lay-out is gekozen of om een uniforme item- en toetsanalyse uit te kunnen voeren. In sommige gevallen is die uniformiteit weer nodig ten behoeve van de gekozen cesuurbepalingsmethode.

Conclusie

Uit onderzoek en ervaring blijkt dat het optimale aantal alternatieven voor gesloten toetsvragen verschilt per toets(domein). Wat de optimale keuze is, is afhankelijk van de situatie, het doel en/of de inhoud van de vraag bepalen het aantal alternatieven. Een goede richtlijn is bij elke vraag zoveel aannemelijke alternatieven te ontwik-

kelen als mogelijk is, meestal drie of vier. Het aantal alternatieven mag in een toets variëren. Ervaringen binnen het toetsdomein en gegevens uit de itemanalyse kunnen tot bijstelling leiden. Als het aantal alternatieven binnen een toets constant moet zijn op basis van beleidsoverwegingen, is het advies kritisch naar de vragen te kijken en zelf te bepalen wat wenselijk is. Onderzoek ondersteunt de stelling dat drie alternatieven in veel gevallen voldoende of zelfs optimaal zijn. ■

Meer lezen?

- Van Berkel, H. & Bax, A. (2002). Toetsen in het hoger onderwijs. Houten/Diegem: Bohn Stafleu Van Loghum.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9), 970–977. <https://doi.org/10.1119/1.1374249>
- Downing, S. M. (1992). True-False, Alternate-Choice, and Multiple-Choice Items. *Educational Measurement: Issues and Practice*, 11(3), 27–30. <https://doi.org/10.1119/1.1374249>

Wat is taalkundig juist? a. Ik ben groter dan jij. b. Ik ben groter dan jou. c. Ik ben groter als jij. Bij deze vraag kan een kandidaat door te puzzelen op het juiste antwoord uitkomen: 'dan' komt twee keer voor en 'jij' komt twee keer voor. Beter is in dit specifieke geval een vierde antwoord toe te voegen: d. Ik ben groter als jou.	Wat gebeurt er met de grafiek $y = ax + b$ als b groter wordt? De lijn verplaatst ... a. naar links. b. naar rechts. c. omhoog. d. omlaag.
---	--

Figuur. Twee voorbeelden van een vraag waarbij beter vier dan drie alternatieven gebruikt kunnen worden.

- org/10.1111/j.1745-3992.1992.tb00248.x
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–333. https://doi.org/10.1207/S15324818AME1503_5
- Mayenga, C. (2009). Mapping item writing tasks on the item writing ability scale. In *Canadian Society of Safety Engineering, XXXVIIth Annual Conference*. Carleton University, Ottawa, Canada. Retrieved from <http://lib-ocs.lib.sfu.ca:8087/fedcan/index.php/csse2009/csse2009/paper/view-File/1966/625>
- Moelands, H. A., Noijons, J., & Rem, J. (1992). *Toetsen met gesloten vragen: een handleiding voor het construeren van toetsen met meerkeuzevragen*. CITO.
- Paes, E. C., & Cate, O. ten. (2009). Meerkeuzevragen met drie, vier of vijf alternatieven: wat is beter? *Tijdschrift voor Medisch Onderwijs*, 28(3), 124–129. <https://doi.org/10.1007/BF03081770>
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3–13. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>

Gesignaleerd

Maximaal zeven jaar celstraf voor fraude examen

Je examen niet op eigen kracht halen, is in China niet aan te raden. Door een wijziging van sectie 284 in het Chinese strafrecht kan je dat nu een maximale celstraf van zeven jaar opleveren. Dat meldt CCTV America.

Vanaf 1 november staat in de wet dat mensen die de boel bedriegen drie jaar celstraf krijgen, terwijl in ernstiger gevallen zelfs een straf tussen de drie en zeven jaar opgelegd kan worden.

<https://www.metronieuws.nl/buitenland/2015/10/maximaal-zeven-jaar-celstraf-voor-fraude-examen>