



**Omdat cesuren
verschillende besluiten,
met verschillende
consequenties
meebrengen,
appelleert de keuze
ervan vaak sterk aan
het rechtvaardigheids-
gevoel.**

Cesurbepaling doe je samen!

De menselijke factor in cesurbepaling

■ **Wilco Emons, Sebastiaan de Klerk & Emmelien van der Scheer**

Dr. W. Emons, dr. S. de Klerk en dr. E. van der Scheer zijn onderzoekers binnen het team Onderwijskundig Onderzoek van Stichting Cito, divisie Onderzoek, Kennis & Innovatie. E-mail: wilcohmemons@gmail.com, sebastiaan.deklerk@cito.nl of emmelien.vanderscheer@cito.nl

Een van de lastigste aspecten van summatieve toetsing is misschien wel het beantwoorden van de vraag: wanneer is het resultaat voldoende? Het kiezen van een slaag-zakgrens, ook wel cesuur genoemd, heeft immers vaak directe – en soms verregaande – consequenties voor de studenten¹. De standaardcesuur van 55% goed of meer is geslaagd geeft lang niet altijd een bevredigend resultaat, zoals de redactionele bijdrage van 2 februari 2022² overtuigend laat zien; maar zie bijvoorbeeld ook Soeting (2017). Ook uit een survey die Stichting Cito via Teacher Tapp³ heeft mogen uitzetten bleek dat met name cesurbepaling lastig gevonden wordt. Daar staat tegenover dat er in de afgelopen decennia diverse methoden zijn ontwikkeld waarmee op een meer systematische manier cesuren bepaald kunnen worden. Deze staan beschreven in toegankelijke literatuur; zie bijvoorbeeld Sanders (2017) en Van Berkel en Wijnen (2014), en we zien het ook regelmatig terug komen in handleidingen voor docenten. De vraag is dan: wat maakt het zo lastig? Is het de complexiteit van de procedures, of spelen andere factoren een rol? In deze bijdrage willen we een bijzonder – en wellicht onderbelicht – aspect van cesurbepaling eruit lichten, namelijk het rechtvaardigheidsgevoel dat wellicht op de achtergrond meespeelt. We doen dit door de gokcorrectie bij meerkeuzevragen en de keuze tussen relatieve en absolute normeringen onder de loep te nemen. We sluiten af met enkele suggesties.

Wat is een cesuur?

Strikt genomen is de cesuur een score op een toets die leidt tot verschillende besluiten over studenten (zie Eggen in Heij, 2016). Je kunt een cesuur dus niet los zien van een besluit, zoals het wel of niet toekennen van studiepunten

of het verstrekken van een certificaat. Meestal wordt het begrip ‘cesuur’ gekoppeld aan de score die leidt tot het cijfer 5,5 – of 6 als alleen met hele cijfers wordt gewerkt – maar die definitie is eigenlijk te beperkt. Denk bijvoorbeeld aan plaatsingsbeslissingen in het voortgezet

¹Daar waar student staat kan ook leerling, cursist, kandidaat worden gelezen.

²<https://www.e-xamens.nl/public/files/Cesuur%20-%20artikel.pdf>

³<https://nl.teachertapp.com/januari-2022/> (8 januari 2022)

onderwijs, waarbij brugklasleerlingen uit een havo/VWO klas gemiddeld een 7 of hoger moeten hebben om naar het VWO te mogen. Hier ligt de cesuur voor toelating bij het cijfer 7.

Omdat cesuren verschillende besluiten met zich meebrengen, met verschillende consequenties, appelleert de keuze ervan vaak sterk aan het rechtvaardigheidsgevoel. Ook al doen we alles volgens vooraf beschreven protocollen, uiteindelijk ligt ook de vraag voor of de gekozen cesuur rechtvaardig is.⁴ Het antwoord op deze vraag laat zich lastig vangen in formele procedures. Het is daarnaast afhankelijk van de context waarbinnen de toetsing plaatsvindt. Zijn bijvoorbeeld de beslissingen onomkeerbaar, of hebben studenten (eindeloze) herkansingsmogelijkheden? En wanneer het gaat om certificering voor risicovolle beroepen zal men weer andere afwegingen maken dan bij low-stakes examens. Zie bijvoorbeeld de literatuur over cesuurbepaling in het medisch onderwijs (e.g., Cohen-Schotanus et al., 2005; Cohen-Schotanus et al., 1996; Ter Burg et al., 2001). Daarbij is het belangrijk om rechtvaardigheid vanuit verschillende perspectieven te bekijken waaronder die van de individuele student, de opleiding en/of beroepsgroep of zelfs vanuit maatschappelijk opzicht.

Wel of niet corrigeren voor gokken?

Gokken is inherent aan meerkeuzetoetsen, maar een goed gegokt antwoord is geen valide weergave van wat iemand kent of kan. Bij meerkeuzetoetsen wordt dan ook vaak geadviseerd om een gokcorrectie toe te passen. Hiermee wil je voorkomen dat studenten door geluk een examen halen. Deze correctie werkt als volgt: Stel je hebt een toets met 40 vierkeuzevragen en je wilt toetsen of studenten over minstens 50% 'ware' kennis beschikken. Zou je niet corrigeren voor gokken, dan zou je de cesuur bij 20 vragen goed of meer leggen. Zou je wél corrigeren voor gokken, dan

komt de cesuur bij 25 te liggen.⁵ De rationale achter deze correctie is als volgt. Als studenten alle vragen blind gokken dan hebben zij naar verwachting 10 vragen goed. Van de overige 30 vragen moet men vervolgens minstens de helft goed hebben – wat neer komt op minstens 50% ware kennis – om te slagen.

Deze rationale klinkt plausibel, maar studenten ervaren deze correctie vaak als onnodig streng.⁶ Hoe rechtvaardig is deze correctie? Om hier meer grip op te krijgen bekijken we een alternatieve methode om gokken te verdisconteren, zoals beschreven in Drenth en Sijsma (2006, p. 156-159). Hierbij wordt niet de cesuur aangepast, maar worden de scores van de studenten zelf naar beneden bijgesteld. Op basis van die bijgestelde scores en de oorspronkelijke cesuur (20 in het voorbeeld) worden beslissingen genomen. De voor-gokken gecorrigeerde scores worden als volgt uitgerekend

$$\text{score_gecorrigeerd} = \# \text{goed} - \# \text{fout} \times 1 / (\text{aantal alternatieven} - 1).$$

Een student die 22 van de 40 vragen goed heeft, komt uit op een gecorrigeerde score van $22 - 18 / 3 = 16$, en heeft het examen niet gehaald.⁷ Het idee is dat de gecorrigeerde score iemands ware kennis beter weergeeft dan het aantal goed. Bovenstaande gokcorrectie is gebaseerd op de volgende uitgangspunten. Er zijn slechts twee mogelijkheden waarop een student tot een antwoord komt: of de student weet het goede antwoord en die geeft hij/zij ook, of hij/zij weet het niet en gokt blind. De consequentie van deze redenering is dat foutieve antwoorden altijd het gevolg zijn geweest van een gok. Het aantal foute antwoorden wordt dus gezien als een indicatie van de mate van gokken. En met vier antwoordalternatieven, verwacht je dat tegenover elke drie fout gegokte antwoorden er waarschijnlijk ook één goed gegokt antwoord

⁴Je zou ook kunnen spreken over 'billijkheid'

⁵De exacte berekening gaat als volgt: $40 \times 1/4 + (40 - 40 \times 1/4) \times 0,5 = 10 + 15 = 25$.

⁶Zie bijvoorbeeld [Gokcorrectie - Scholieren.com forum](https://www.gokcorrectie-scholieren.com/forum)

⁷Voor goed-foutvragen is deze gecorrigeerde score gelijk aan het aantal goed min het aantal fout

staat. Dit verwachte aantal goed gegokte antwoorden wordt van het aantal goed afgetrokken en dit levert de gecorrigeerde score op.

Beide methoden vertrekken vanuit hun eigen uitgangspunten, maar rekenen we vervolgens de scores via een lineaire transformatie om naar cijfers⁸ – scores zijn op zichzelf meestal niet heel informatief – dan blijkt dat het niet uitmaakt welke van de twee methoden wordt gebruikt. Neem een student die 30 vragen goed heeft. Zou je de gecorrigeerde cesuur gebruiken en de score omrekenen naar een cijfer volgens de lineaire methode dan krijg je $10 - (40 - 30) \times 4.5 / (40 - 25) = 7$. Zou je de gecorrigeerde score gebruiken, en de oorspronkelijke cesuur, dan krijg je $10 - (40 - 26.7) \times 4.5 / (40 - 20) = 7$. Met andere woorden, beide correctiemethoden leiden tot dezelfde beslissingen en waarderings. Door de gokcorrectie vanuit verschillende perspectieven te benaderen zien we dat beide vormen van gokcorrectie op vrij restrictieve aannames gebaseerd zijn. Want hoe aannemelijk is het dat studenten altijd blind gokken als ze het antwoord niet weten? En als er wordt gegokt, hebben alle alternatieven een even grote kans of worden op voorhand enkele foute alternatieven uitgesloten? Maar het belangrijkste is misschien nog wel de vraag of een student die oprecht een fout antwoord geeft omdat hij/zij de stof niet beheerst door de gokcorrectie niet extra wordt gestraft? Wellicht verklaart dit het onbehagen van de studenten.

Daarbij dient men zich te realiseren dat het wel of niet toepassen van een gokcorrectie op de cesuur niet alleen de slaag-zakbeslissing beïnvloedt, maar ook de waardering die aan prestaties wordt gegeven. Neem opnieuw het voorbeeld van een toets met 40 vierkeuzevragen. Zou je de cesuur op “50% ware kennis” zetten, dan zou een student met 30 vragen goed een 7,8 halen indien geen gokcorrectie wordt toegepast, en een 7,0 als er wel gecorrigeerd is voor gokken. Het verschil is bijna 1

volledig cijferpunt.

Tegelijkertijd wil je ook niet dat studenten simpelweg door te gokken het examen onterecht halen. Maar hoe groot zijn die kansen eigenlijk? Stel dat een student een examen maakt van 20 vier-keuzevragen, waarbij de cesuur bij 10 of meer ligt, en daarbij *alles* gokt. In dit geval is de kans op slagen 1.4%. Dus je moet als onwetende student redelijk veel geluk hebben om de toets te halen. Met 40 vragen en vier alternatieven, een cesuur van 20, is de kans dat hij/zij slaagt nog véél kleiner, namelijk 0.057% (ongeveer 6 op 10.000). Ook het aantal alternatieven, en de aantrekkelijkheid van alternatieven, spelen een rol. Bij goed-foutvragen heb je 50% kans op goed gokken. Als de cesuur dan ook nog eens bij 50% goed of meer geslaagd ligt, dan wordt gokken wel heel lucratief. De gebruikelijke gokcorrectie bij goed-fout vragen is echter weer erg streng omdat studenten ook echt fouten maken. Gokken is dus vooral een probleem bij korte toetsen en een beperkt aantal alternatieven. Wanneer men de slaagkans door gokken wil verkleinen – helemaal voorkomen is onmogelijk – dan is het beter om meer vragen te stellen, en/of vragen met meerdere alternatieven aan te bieden dan te vertrouwen op een gokcorrectie. Hiermee vergroot je tegelijkertijd de betrouwbaarheid van de toets, en daarmee de betrouwbaarheid van de beslissingen die op basis van de toets genomen worden.

Relatief of absoluut?

Het is gebruikelijk om een onderscheid te maken tussen relatieve normen en absolute normen (zie bijvoorbeeld Van Berkel & Wijnen, 2014). Bij een relatieve norm hangt het resultaat van een individuele student af van de prestaties van andere studenten die de toets maken. Van een relatieve norm is bijvoorbeeld sprake als de cesuur bij een vast slagingspercentage (bijv. 60%) wordt gelegd. De rationale is dat de gemiddelde vaardigheid van studentenpopulaties over tijd stabiel blijft. Bij een absolute cesuur wordt van tevoren een

⁸Voor scores boven de cesuur is de lineaire omzetting gegeven door: $10 - (\text{maxscore} - \# \text{goed}) \times 4.5 / (\text{maxscore} - \text{cesuur})$ (e.g., Stichting Cito, 2020). Zie ook [Omzettingstabel \(cito.nl\)](https://www.cito.nl/omzettingstabel)

cesuur vastgesteld. Bijvoorbeeld, de student is geslaagd bij 50% of meer van de vragen goed. De uitkomst voor een individuele student is onafhankelijk van wat zijn of haar medestudenten doen. Deze aanpak gaat er vanuit dat de moeilijkheid van de examens constant zijn en adequaat zijn afgestemd op de gewenste inhoudelijke standaard. Er zijn ook methodes die relatieve informatie combineren met een absolute cesuur, zoals de methode van Cohen-Schotanus en Van der Vleuten (2010). De keuze voor absoluut of relatief raakt opnieuw aan

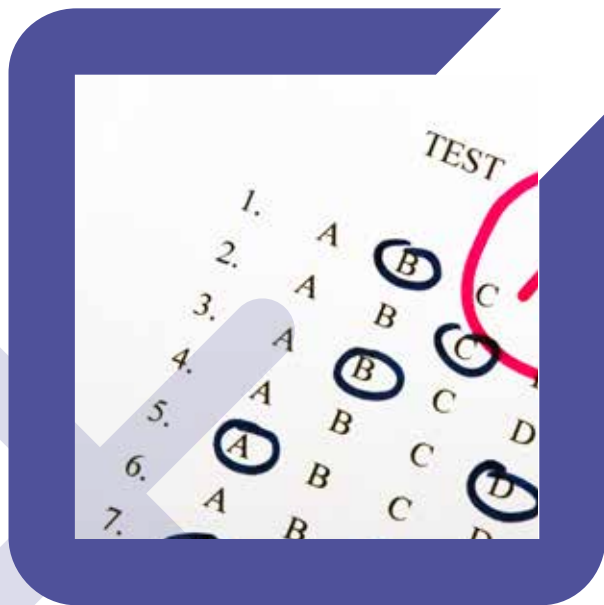
de vraag naar rechtvaardigheid. Want is het rechtvaardig dat een individuele student voor het slagen of zakken afhankelijk is van hoe zijn medestudenten presteren? Tegelijkertijd, is het rechtvaardig om ongeacht hoe goed het onderwijs is verlopen, dezelfde standaarden op te leggen aan studenten? Opnieuw is dit geen kwestie van goed of fout, maar wel één van weloverwogen en fundamentele keuzes.

Cesuur bepalen doe je samen en in samenhang!

De belangrijkste constatering is misschien wel dat dé juiste cesuur niet bestaat. Wat wel bestaat zijn weloverwogen, op onderzoek gebaseerde, transparante, geloofwaardige en collegiaal gedragen cesuren (Cohen-Schotanus & Van der Vleuten, 2010). Intensief overleg tussen opleiders, docenten, belanghebbenden, waarbij alle facetten – ook de menselijke factor – aan de orde komen is cruciaal. Wellicht is op het laatste punt nog winst te behalen (zie bijvoorbeeld Emons et al., 2021). Standaardbepalingsmethodes zoals die van Angoff (e.g., Sanders, 2017) geven concrete kaders, hoewel vaak tijdrovend en voor veel docenten misschien tamelijk abstract. Het is belangrijk om te blijven zoeken naar praktische handvatten om cesuurbepaling in de dagelijkse praktijk zo goed mogelijk vorm te geven. ■

Referenties

- Cohen-Schotanus, J., Schönrock-Adema, J., & Scherpbier, A. J. J. A. (2005). Welk percentage gezakte studenten is bij een bepaalde toetsvorm acceptabel? De mening van NVMO-congresgangers. *Tijdschrift voor Medisch Onderwijs*, 24, 184-189.
- Cohen-Schotanus, J. & Van der Vleuten, C. P. M. (2010). A standard setting method with the best performing students as the point of reference: Practical and affordable. *Medical Teacher*, 32, 154-160.
- Cohen-Schotanus, J., Vleuten, C.P.M. van der, & Bender, W. (1996). Een betere



Wanneer men ‘slagen door gokken’ wil beperken – want helemaal te voorkomen is onmogelijk – dan lijkt het rechtvaardiger om wat meer vragen te stellen, en meerdere alternatieven aan te bieden.

cesuur bij tentamens. *Onderzoek van Onderwijs*, 25, 54-55.

- Heij, K. (2016). *Waar ligt de cesuur: een masterclass*. Toets! Magazine Bureau ICE <https://ris.utwente.nl/ws/portalfiles/portal/6931825/Waar+ligt+de+cesuur%2C+een+masterclass.pdf>
- Emons, W. H. M., Van der Scheer, E. A., Hemker, B. T. & Veldhuizen, J. A. A. C. (2021). *Cijfers, cijfers, cijfers: Een verkenning naar toetsing in het hoger onderwijs* [onderzoeksrapport]. Stichting Cito.
- Drenth, P. J. D., & Sijtsma, K. (2006). *Testtheorie: Inleiding in de theorie van de psychologische test en zijn toepassingen*. Bohn, Stafleu, & Van Loghum.
- Sanders, P. (2017). Toetsen op school (herziene versie). *Stichting Cito*. https://www.cito.nl/-/media/files/kennis-en-innovatie-onderzoek/toetsen-op-school/cito_toetsen_op_school.pdf?la=nl-nl
- Soeting, J. (2017). Klopt dit wel? 55% van de punten is (een) voldoende. *Examens*, 3, 55-56.
- Stichting Cito (2020). *Beoordelen en becijferen: bepalen van slaag-zakgrens & omzetten van score naar cijfer* [whitepaper]. https://www.cito.nl/-/media/files/kennis-en-innovatie-kennisplein/cito_whitepaper_bepalen_slaag_zakgrens_omzetten_scores_cijfers_tiptoa.pdf?la=nl-nl&hash=4B9F037AC51081FA0FFB092F1F7B2CE720408E98
- Ter Burg, J., Emons, W. H. M., Mol, S. S. L. & Schuwirth, L. W. T. (2001). Bepaling van de referentiewaarde bij screenende kennistoetsen door het NHG met de methode van Angoff. *Tijdschrift voor Medisch Onderwijs*, 20, 17-24.
- Van Berkel, H. & Wijnen, W. (2014). Normeren en cijfergeven. In Van Berkel, Bax, Joosten-ten Brinke (Eds.). *Toetsen in het hoger onderwijs* [derde editie]. Bohn, Stafleu & Van Loghum.

Gesignaleerd

Fraude met valse diploma's in zorg neemt toe: cashen over de rug van kwetsbare cliënten

18 juli 2022 16:42 RTL nieuws

Het aantal personeelsleden in de zorg dat is betrapt met een vals diploma is flink gestegen. Het Openbaar Ministerie (OM) en zorgmedewerkers slaan alarm. Diplomafraudeurs geven medicijnen of leggen een infuus aan, terwijl ze daar helemaal niet voor zijn opgeleid. "Het gaat om ernstige zaken vanwege mogelijke gezondheidsrisico's voor cliënten", stelt het OM. In het eerste half jaar van 2022 ontving de Inspectie Gezondheidszorg en Jeugd (IGJ) 59 meldingen van fraude met valse diploma's en Verklaringen Omtrent het Gedrag (VOG). Dat blijkt uit cijfers die de onderzoeksredactie van RTL Nieuws heeft opgevraagd. Dat zijn nagenoeg evenveel meldingen als de voorbije jaren in een heel jaar.

"We zien mensen die niet het juiste diploma hebben, geen werkervaring of soms een nep-diploma. Valse diploma's hebben we niet eerder gezien of opgemerkt. Dat is echt van het afgelopen jaar."

Antwoorden op kamervragen hierover zijn na te lezen op: <https://www.rijksoverheid.nl/documenten/kamerstukken/2022/08/26/antwoorden-op-kamervragen-over-het-bericht-fraude-met-valse-diplomas-in-zorg-neemt-toe-cashen-over-de-rug-van-kwetsbare-patienten>