

Cesuur

Een lastig toetsaspect bij vooral summatieve toetsen

Van de redactie van Examens

2 februari 2022

Uit het beoordelen van enkele honderden summatieve tentamens en examens (hierna te noemen toetsen) uit vooral het hbo, maar ook vanuit het mbo, komt naar voren dat het bepalen van de cesuur (grens tussen zakken of slagen) en het normeren (omzetten van een score op een toets naar een cijfer of een andere waardering) als lastig wordt ervaren. Evenals overigens het psychometrisch analyseren van de resultaten op toetsen op basis van p-, a-, Rit- en betrouwbaarheidswaarden. In het algemeen kan gezegd worden dat de getalsmatige kant van toetsen lastig wordt bevonden. In deze bijdrage aan deze E-Nieuwsbrief wordt ingegaan op de cesuur en cesuurmethoden voor enkele veel voorkomende toetsvormen.

Met een summatieve toets wordt een zak-/slaagbeslissing genomen op basis van de behaalde score van de student op de toets. Aan deze beslissing is vaak een zwaarwegende consequentie verbonden. Door te slagen mag de student bijvoorbeeld een auto besturen, doorgaan naar het volgende studiejaar of zelfs een beroep gaan uitoefenen. Door te zakken mag dit allemaal ineens (nog) niet. De zak-/slaagbeslissing dient dus zorgvuldig en zo objectief mogelijk te worden bepaald. Zorgvuldigheid en objectiviteit hebben vooral te maken met de transparantie, validiteit en de betrouwbaarheid van de meting die leidt tot de behaalde score van de student.

Naast dat de score van de student bepaald wordt, moet worden vastgesteld waar de grens tussen zakken en slagen zich precies bevindt. Dus bij welke score op een toets is de student geslaagd. Dit wordt aangeduid met het begrip cesuur (en niet censuur zoals soms wordt geschreven).

Cesuur is geen cijfer

De cesuur wordt uitgedrukt in een score op de toets en is dus geen cijfer. Wel is het zo dat in Nederland bij het cijfer vanaf een 5,5 een student geslaagd is. De cesuur is dan ook de score op de toets die leidt tot het cijfer 5,5, of een 6 als wordt gewerkt met hele cijfers of tot een andere waardering.

Wat is cesuur?

Wat cesuur is, is heel duidelijk namelijk de grens tussen een onvoldoende of een voldoende, tussen zakken of slagen voor een toets. Maar hoe bepaal je dan die grens? In Nederland wordt breed gebruik gemaakt en is het algemeen geaccepteerd dat de student iets meer dan de helft van de maximale score op de toets moet behalen om te slagen en dit is dan 55%. Dit is een manier van absoluut normeren. Daarnaast bestaan ook methoden voor relatief en compromisgericht normeren, maar daar zal in dit artikel niet op worden ingegaan. worden.

De Inspectie van het Onderwijs heeft de 'Normenbundel Exameninstrumenten Toezichtkader 2012, versie januari 2014' (voor het mbo) uitgebracht. Deze Normenbundel zegt over de cesuur het volgende: *'De cesuur ofwel de zak/slaaggrens van elke toets, ongeacht de toetsvorm (dus ook de BPV (red: BeroepsPraktijkVorming bij een leerbedrijf)) ligt op minimaal 55%.'* Dat ongeacht de toetsvorm overigens niet altijd klopt wordt hier later nog toegelicht.

Belangrijk is dat de Inspectie spreekt over minimaal 55%. De Inspectie benoemd ook het volgende:

'Er is een verantwoorde procedure gevolgd voor het bepalen van de cesuur, zowel voor de examenonderdelen als voor het examen als geheel. De cesuur van het exameninstrumentarium ligt op het niveau waarop de student voldoet aan de vereisten. De cesuur is zodanig opgesteld dat de beoordeling (en daarmee de waardering van de kennis, vaardigheden en houding van de student) in totaal niet lager uitpakt dan voor de vakbekwaamheid voor het beroep noodzakelijk is. Er is sprake van een evenwichtige, niet oneigenlijke puntentoekenning die recht doet aan het belang van eindtermen of kerntaken en werkprocessen. Er is sprake van een weging van de examenopdrachten/-toetsen onderling, die geen afbreuk doet aan de genoemde cesuur.'

Een farmaceut die zogeheten artsenbezoekers inzet om bijvoorbeeld knieprothesen te adviseren aan specialisten, eist dat de artsenbezoekers voor een interne toets met gesloten vragen moeten slagen met een 85% score. De artsenbezoekers zijn veelal afgestudeerde artsen, maar die over het specialisme knie net zoveel en liever nog meer moeten weten als de specialist zelf (zie de verantwoorde procedure uit de tekst van de Normenbundel). Pas bij 85% voldoet de (aankomend of nog steeds aantoonbaar bekwame) artsenbezoeker aan het vereiste. Dit is vastgesteld door de farmaceut als noodzakelijk voor de vakbekwaamheid. En met de 85% wordt waarschijnlijk recht gedaan aan de dekking van de achterliggende toetsmatrijs met de leerdoelen en de beheersingsniveaus zoals kennis, inzicht, toepassen, analyse, evalueren en creëren.

De Inspectie beschrijft exact waar het om gaat: een cesuur komt verantwoord tot stand, ligt op het niveau waarvoor de student wordt opgeleid en doet samen met de toets als geheel recht aan de eindtermen.

Wordt uitgegaan van de standaardcesuur van 55% dan is dat eigenlijk de kort door de bocht-methode waarbij we ervan uitgaan dat de eindtermen transparant, valide en betrouwbaar worden beoordeeld. Dat we in het reguliere onderwijs deze methode veel gebruiken komt doordat we uitgaan van de nominale student. Dat is de student die, als hij altijd slaagt met een 55% score, slaagt voor de opleiding en zijn opleidingsduur is ook nominaal, dus binnen de tijd die ervoor staat al dan niet met succesvolle herkansingen binnen de gestelde termijnen.

Het hele reguliere onderwijs is dus ingericht op de zesjes-student. Een eigenlijk kunnen we dus daarom die 55% cesuur binnen het reguliere onderwijs goed hanteren. Ook de Inspectie zegt: 'Ga uit van minimaal 55%', maar meer dan 55% is eigenlijk niet meer nominaal. Dus verantwoord dan waarom de cesuur meer moet zijn. Denk bijvoorbeeld aan medisch rekenen. Sommigen ziekenhuizen eisen een 100% goedscore omdat het risico van een verkeerde toegediende dosering dodelijk kan zijn. Dus dat een verpleegkundeopleiding voor het examen medisch rekenen bijvoorbeeld een cesuur van 80% eist is verklaarbaar en past bij de vereiste vakbekwaamheid waaraan de student in dat vak moet voldoen. En eigenlijk is dat voor dat beroep weer nominaal; of hoort het dat te zijn.

Kort samengevat: ga binnen het reguliere onderwijs uit van 55% en verklaar waarom een cesuur soms een hoger percentage moet hebben.

Voor het op basis van de vereiste bekwaamheid bepalen van de cesuur, dus niet via de 55%-regel, zijn verschillende zogenaamde standaardbepalingen beschikbaar zoals die van Angoff, Nedelsky, Ebel, Jaeger, Idea of Kernitem. We laten deze hier nu buiten beschouwing, vanwege de bewerkelijkheid en de uitgebreidheid hiervan om ze toe te lichten. Wellicht in een volgende E-nieuwsbrief.

Wat we ook buiten beschouwing laten is de relatieve cesuurmethode waarbij achteraf op basis van de gemiddelde score op de toets, gemaakt door een groep studenten, de cesuur wordt berekend (methode Wijnen bijvoorbeeld). Ook laten we de compromismethode buiten beschouwing (methodes van Hofstee, Beuk en Cohen-Schotanus bijvoorbeeld). Wellicht in een volgende E-nieuwsbrief.

De 55%-cesuurregel is typisch een vorm van een absolute cesuur die vooraf, dus voor afname van de toets, wordt opgesteld en gedeeld wordt met de studenten. Dus ze weten vooraf precies wanneer ze geslaagd zijn voor de toets. Overigens wordt dit niet altijd vooraf gedeeld met de studenten wat vanuit het oogpunt van transparantie wel zou moeten.

Cesuur per toetsvorm toegelicht

Gelet op het voorgaande gaan we voor het gemak uit van een 55% cesuur tenzij anders van toepassing zoals bij opdrachtvormen het geval kan zijn.

Toets met gesloten vragen

Bij een toets met gesloten vragen moet de afweging worden gemaakt of wel of niet rekening gehouden wordt met de blinde raadkans. Door blind te raden kan een student bij een toets met bijvoorbeeld 40 vragen met elk 4 alternatieven (blinde raadkans is 25%) en 1 punt per goed antwoord, statisch gezien 10 punten scoren door uitsluitend te raden. Deze 10 punten geven niet de bekwaamheid van de student weer. Vandaar dat vaak ervoor wordt gekozen om de blinde raadkans te verdisconteren in de cesuur. Deze worden dan hoger dan de 55% van 40 punten (= 22 punten), namelijk 26 punten. Dit wordt als volgt berekend:

$10 + 0,55 \times (40 - 10) = 26,5$ en dat wordt afgerond naar 26 punten vanwege de meetfout die de toets sowieso heeft en doen we dat studentgericht. Bij de cesuur van 26 punten hoort dan het cijfer 5,5.

De berekening is gebaseerd op de aanname dat door blind raden statistische gezien 10 punten kunnen worden behaald en dat de kandidaat van wat overblijft (40 – 10) 55% goed moet hebben.

Bij een toets van bijvoorbeeld 75 vragen die alleen met juist/onjuist beantwoord kunnen worden (50% blinde raadkans is 37,5 punten) komt de cesuur uit op: $37,5 + 0,55 \times (75 - 37,5) = 50,4 = 50$ punten. Dus pas bij 50 vragen goed krijgt de student een 5,5. Zonder rekening te houden met de blinde raadkans zou de student bij 41 punten zijn geslaagd.

De keuze om wel of geen rekening te houden met de blinde raadkans heeft nogal wat invloed op het juist zakken of net slagen van een student en dus op de mogelijke consequenties. Hetzelfde geldt natuurlijk voor de keuze om uit te gaan van 55% als cesuurpercentage.

Binnen het hbo kwam een keer een toets met gesloten vragen voor met een cesuurpercentage van 50%, zonder rekening te houden met de blinde raadkans. In relatie tot de twee voorbeeldtoetsen van hiervoor zijn de studenten dan geslaagd bij respectievelijk 20 en 37 punten in plaats van de 26 en 50 punten. En nog vonden de studenten het moeilijk om te slagen voor de toets!

Bij gebruik van een digitaal toetsstelsel zoals TestVision, RemindoToets of ANS kunnen ook andere vraagvormen dan de meerkeuzevraagvorm worden gebruikt. Denk bijvoorbeeld aan de meervoudig-juist-vraag (bijvoorbeeld 3 van de 7 antwoordopties zijn goed) of een sorteervraag (zet de antwoordopties in de juiste volgorde). In dat geval is het mogelijk om een antwoord gedeeltelijk goed te rekenen; dus een score ergens tussen 0 en 1 punt. Ook hier is sprake van een blinde raadkans. De

gedeeltelijke score en raadkans zijn complex om te berekenen, maar daar zorgen de toetssystemen voor, onder meer op basis van de product- en quotiëntregel (zie Hans Vos, Marlies Kloppenburg en Onno Tomson in Examens 2010-3 over Optimale uniforme scoringsregels voor innovatieve vraagvormen). De cesuur instellen op 55% en rekening laten houden met de blinde raadkans en het toetssysteem bepaalt vervolgens bij hoeveel scorepunten de 55%-cesuur komt te liggen.

Toets met open vragen

De cesuurberekening is hier eenvoudig en er is geen sprake van blind raden. De cesuur bij 55% is dan 55% van het aantal punten dat maximaal kan worden gescoord op de toets door het goed beantwoorden van alle vragen. Stel dat 60 punten kan worden behaald met bijvoorbeeld 12 vragen, dan is de cesuur 33 punten (55% van 60).

Toets op basis van opdrachten

Toetsen op basis van opdrachten worden vaak beoordeeld met:

- integratieve beoordelingscriteria waarbij meerdere leeruitkomsten/leerdoelen in samenhang met elkaar beoordeeld worden op basis van een rubric met drie of meer beoordelingsniveaus en indicatoren of met de steeds meer gebruikt single-point-rubric.
- analytische beoordelingscriteria die per beoordelingscriterium de leeruitkomsten/leerdoelen sterk volgen waarbij de beoordeling plaatsvindt op een ratingscale met drie of meer beoordelingsniveaus zonder het gebruik van indicatoren.

Bij het gebruik van een rubric wordt per beoordelingscriterium met zogenaamde indicatoren precies beschreven wanneer op welk niveau daaraan wordt voldaan. Dus wanneer is er voldoende aan voldaan, wanneer is het goed, of zelfs uitstekend en wanneer is het onvoldoende (zoals bij een vierpuntschaal: O/V/G/U). Bij het gebruik van een single-point-rubric wordt alleen uitgegaan van het niveau voldoende en wordt toegelicht waarom iets niet voldoende is of ruim meer dan voldoende is. Bij een ratingscale wordt in algemene termen beschreven dat een voldoende overeenkomt met de geldende standaard, dat dat bij een onvoldoende daar onder zit en bij een goed of nog hoger, dat aan meer dan de standaard wordt voldaan.

Veel gebruikte niveauschalen bij een rubric of een ratingscale kennen drie, vier of vijf niveaus. Meer dan dit aantal niveaus levert al snel problemen bij het kiezen van het best passende niveau. De niveaus worden bij summatieve opdrachten meestal aangeduid met O/V/G of O/V/G/U of ZO/O/V/G/U. Vaak worden hieraan scorepunten gekoppeld zoals respectievelijk 1/2/3, 0/1/2/3 of 0/1/2/3/4 punten. Of zoals bij de laatste (ZO/O/V/G/U) kom je ook wel 2/4/6/8/10 punten tegen. Niet voor niets krijgt de V dan de waarde 6 toegekend.

Hierna volgen enkele voorbeelden van berekeningen om tot een cesuur te komen bij het gebruik van een rubric of een rating scale; dit maakt voor de berekening niets uit en is voor beide hetzelfde.

Stel dat sprake is van 12 beoordelingscriteria en dat een driepuntsschaal (O/V/G met 1/2/3 punten) wordt gebruikt. Dit geldt zowel bij het gebruik van een rubric als van een ratingscale.

Cesuur bepalen met de alles-V-regel (alles voldoende)

De meest recht-toe-recht-aan-methode is dat de student geslaagd is bij alles V (overigens is dit al de eigenlijke cesuurkeuze). De cesuur in scorepunten is dan $12 \times 2 = 24$ punten. De maximale score is $12 \times 3 = 36$ punten en de minimale score is 12 punten (12×1). Qua cijfers betekent dit bij 12 punten een 1, bij 24 punten een 6 en bij 36 punten een 10. Tusseliggende scores worden lineair verdeeld tussen de 1 en de 6 en lineair tussen de 6 en de 10. Eenvoudig dus.

Maar bij deze berekening is sprake van zogenaamde interne compensatie. Twee onvoldoendes kunnen worden gecompenseerd door tweemaal een G. Of zelfs zes onvoldoendes door zes maal een G. De vraag is of zoiets wenselijk is. Vaak wordt per beoordelingscriterium gekeken of een dergelijk criterium niet kritisch is; mag niet onvoldoende worden gescoord. De cesuur blijft dan 12 punten, maar er wordt aangegeven dat bijvoorbeeld criterium 4 en 6 niet met een O mogen zijn beoordeeld en/of dat er bijvoorbeeld maar driemaal een O mag worden gescoord. Anders is de totaalscore sowieso onvoldoende. De eenduidige cesuur wordt dan omgeven met randvoorwaarden.

Maar als sprake is van een cijfer én er is sprake van een of meer randvoorwaarden wat leidt tot een onvoldoende hoe bepaal je dan het cijfer? Want de student kan 11 keer een goed hebben gescoord maar op een knock-outcriterium een O, dus is de totaalscore onvoldoende, ondanks de 33 gescoorde punten. Daar kun je het bij laten, maar de student weet dan niet hoe erg zijn zakken is uitgedrukt in een cijfer. Daar zijn rekenregels voor bedacht zoals bijvoorbeeld de volgende: deel 5,4 (het hoogste onvoldoende cijfer) door het aantal criteria, in dit geval $12 = 0,45$. Voor elk voldoende of goed gescoord criterium ontvangt de student 0,45 cijferpunt. Dus in het voorbeeld krijgt de student het cijfer 4,9 ($11 \times 0,45$).

Cesuur bepalen met de 55%-regel

Een geziene methode die fout is, is de volgende; hier gaat de Normenbundel van de Inspectie ook de mist in (cesuur ligt op minimaal 55% ongeacht de toetsvorm). De cesuur wordt berekend door 55% te nemen van de maximale score; op basis van ons voorbeeld $36 \text{ punten} = 19,8 = 19$ of 20 punten (doe maar 19 vanuit het aspect van beoordelingsfouten waarvan we de student niet de dupe laten worden). Tot zover lijkt er nog niets aan de hand te zijn. Behalve dat de student voor dezelfde opdracht niet met 24 punten (zie de V-methode van hiervoor), maar met 19 punten al slaagt.

In het voorbeeld wordt uitgegaan van een driepuntsschaal met O/V/G met 1/2/3 punten. Stel dat de beoordelingsschaal uitgaat van 0/1/2 punten dan ontstaat het volgende. De cesuur bij 12 maal V is 12 punten en de student slaagt dus als hij 12 punten of meer behaalt. De minimale score is 0 punten en de maximale score is 24 punten. De cijferverdeling is weer een 1 bij 0 punten, een 6 bij 12 punten en een 10 bij 24 punten. Voorts weer twee lineaire verdelingen voor de cijfers tussen 1 en 6 en tussen 6 en 10.

Bij de 55% regel is de student geslaagd bij 13,2 punten (55% van 24) = 13 punten. Volgens de V-regel slaagt de student met 12 punten. De vraag is waarom moet een student meer punten behalen om te slagen dan dat hij alles op het niveau van voldoende heeft aangetoond; uitgaande van de nominale student en wat de 55%-regel ook impliceert. Bovendien is deze rekenmethode gevoelig voor de scorepunten op de schaalverdeling.

Cesuur bepalen met de 75%-V-regel

Wat ook voorkomt bij opdrachten is dat de cesuur minimaal 75% van de beoordelingscriteria is die daarbij minimaal als voldoende moeten zijn beoordeeld. In het voorbeeld dus 75% van minimaal 24 punten (= 18 punten) of 75% van minimaal 12 punten (9 punten) in het voorbeeld hiervoor met de andere puntenschaal. Deze 75%-V-regel is geheel legitiem want een cesuur moet passend zijn voor het doel van de toets met het vereiste niveau van vakbekwaamheid. De Inspectie geeft immers aan dat: *'De cesuur van het exameninstrumentarium ligt op het niveau waarop de student voldoet aan de vereisten. De cesuur is zodanig opgesteld dat de beoordeling (en daarmee de waardering van de kennis, vaardigheden en houding van de student) in totaal niet lager uitpakt dan voor de vakbekwaamheid voor het beroep noodzakelijk is.'* Dus als het beroepenveld, de deskundigen en de docenten met elkaar vinden dat daaraan wordt voldaan dan vormen 18/9 punten dus de cesuur.

Concluderend

Samengevat hebben we in een oogwenk voor de cesuurbepaling van opdrachten drie uiteenlopende waarden gezien voor eenzelfde casus, net welke cesuurmethode wordt gebruikt:

- de alles-V-regel met eventueel randvoorwaarden om te sturen ten aanzien van kritische beoordelingscriteria;
- de 55%-regel die geheel verschillende cesuren laat zien vergeleken met de alles-V-regel afhankelijk van de scorepunten per beoordelingsniveau;
- de 75%-regel van de voldoende.

De Inspectie heeft helemaal gelijk dat een cesuur moet passen bij wat voor de vakbekwaamheid voor het beroep noodzakelijk is. Het voorgaande laat zien de drie methodes leiden tot verschillende cesuren. Hier wordt geen recht gedaan aan wat de Inspectie en anderen bedoelen met wat een passende cesuur moet zijn. De cesuur wordt bepaald door de cesuurmethode en eigenlijk zonder te

kijken naar de inhoud en de context van de toets. De student zal zelf het meeste blij zijn met de 75%-V-regel, want daar slaagt de student met de laagste scorepunten, maar of dat de passende cesuur is, is maar de vraag. Het is dus niet de cesuurregel waar het om gaat maar de vraag wat de passende cesuur zou moeten zijn. Kies dus niet gedachteloos een cesuurregel, maar kies bewust. Maar laat er in ieder geval geen willekeur bestaan tussen cesuurregels, daar lijkt het nu wel op.

Note

Er is nagevraagd wat de bron is van de 75%-V-regel. Niemand die daar een goed antwoord op kon geven; *'Doen we al jaren zo'*, *'Dat is toch logisch, je mag ook wel wat fout hebben'* of *'Zoiets hebben we ooit opgetekend van iemand van de Inspectie'*. Niemand die zei dat de 75%-V-regel de minimale vakbekwaamheid omvat die voor het beroep noodzakelijk is. En geldt dat dan voor elk beroep?

Toch maar weer in de Normenbundel van de Inspectie gedoken, die heet niet voor niets Normenbundel. Wat staat daar als wordt gezocht op 75%: *'De instelling voldoet aan de norm als meer dan 75% van de eindtermen, qua inhoud en toetsvorm, juist worden geëxamineerd.'*

Dat minimaal 75% van de eindtermen in een examen terug moeten komen, is iets anders dan dat 75% van de te behalen beoordelingscriteria met een voldoende moeten worden afgesloten.

Wellicht moeten we toch maar eens naar Angoff, Nedelsky, Ebel, Jaeger, Idea of Kernitem gaan kijken?

Het bovenstaande laat zien dat achter een eenvoudige definitie, de cesuur is de scoregrens tussen zakken en slagen, een hele complexe wereld schuil gaat. Het gedachteloos een cesuur hanteren en slaafs de meute volgen, doet geen recht aan de beslissing die we met een toets nemen richting beroepenveld, student en maatschappij.

Graag zien we als redactie van Examens reacties op het artikel tegemoet. Wellicht klopt een en ander niet en/of kunnen we een en ander nog concreter maken.