



HO 5a Vastleggen van de cesuur, cijfers geven en de uitslag bekendmaken

Cesurbepaling en cijfers geven

De cesuur is de zak/slaaggrens van een toets. Bij de cesuur stel je jezelf dus de vraag welke score van de student resulteert in een nèt voldoende (5,5) of een nèt onvoldoende. De cesurbepaling probeer je (zoals aangegeven in HandOut 1b en 1c) zo precies en valide als kan vooraf in te schatten op basis van inhoudelijke argumenten. Inzichtelijkheid in hoe de voldoende grens wordt berekend is van belang in communicatie met studenten. Een simpele navolgbare werkwijze (transparantie) heeft de voorkeur boven ingewikkelde en geavanceerde formules. Voor deze laatste fase in het toetsproces gaan we ervan uit dat de beheersingsstandaard voor een voldoende al in de ontwerpfase bij benadering is ingeschat. Dit noemen we een **absolute cesurbepaling**. Het is verstandig om achteraf op basis van gebleken kwaliteit van de toets naar redelijkheid nog aanpassingen door te voeren, die gevolgen hebben voor zowel de vastgelegde cesuur en de cijfer toekenning. Deze methode van het aanpassen en vaststellen van de cesuur wordt ook wel de **compromis-methode** genoemd. Bij een **relatieve cesurbepaling** wordt in tegenstelling tot de absolute (en compromismethode) de prestatie van de individuele student afgemeten aan de prestatie van de groep.

Bij beslissingen ('high stakes test') die grote consequenties hebben voor de student is het uiteraard van belang dat dat zorgvuldig gebeurt en met een methode die is uit te leggen naar de student en in-, en extern is te verantwoorden. Bij de beslissing gezakt of geslaagd levert de toets dus evidentie over de mate waarin de student voldoet aan de eindvereisten. De keuze voor het referentiepunt is daarbij van doorslaggevend belang: is het referentiepunt inhoudelijk gedefinieerd (verwijzing naar een standaard met behulp van een toetsmatrijs)? Een ander referentiepunt is de 'Face validity' zoals de ervaring en de verwachtingen de docent zelf van deze groep studenten. Stroken de resultaten van de studenten met de indrukken die de docent vooraf had? Een weer heel ander referentiepunt is waar visitatiecommissies naar kijken de geleverde gemiddelde inspanning van de student voor deze cursus in relatie met het slaagpercentage.

De volgende vier mogelijkheden laten verschillende indrukken achter en roepen vragen op, die reden kan zijn om meer precies te kijken naar de oorzaak, het onderwijs, de toets en de gestelde eisen en cesuur.

Slaag%	Studie-inspanning	
	Laag	Hoog
Hoog	Cursus te licht, toets te gemakkelijk?	Loon naar werken
Laag	Cursus wel voldoende uitdagend? Programmering studeerbaar? Parallelvak energieslurper?	Struikelblok. Te hoge eisen? Didactisch te verbeteren?

Cesuur en toekennen van cijfers: in aparte stappen of integraal?

In de onderwijspraktijk is de cesuur (zak-slaaggrens) vaak een resultante van de lijn waarlangs scores in cijfers worden omgezet. Maar, het is goed om de verschillende stappen van meten tot het nemen van beslissingen uit elkaar te houden.



In figuur 1 hiernaast zijn de stappen in het proces onderscheiden naar formatieve deel en het summatieve deel van toetsen.

Als de scores op de vragen zijn toegekend en gesommeerd zijn de eindscores van de studenten vastgesteld.

Nu gaat het er vervolgens om de scores in cijfers weer te geven, waarin de cesuurscore expliciet een rol speelt of impliciet komt vast te staan. Om een en ander inzichtelijk te maken gaan we wat dieper in op de score-cijfer transformatieschaal.

Figuur 1 Stappen in het toetsproces van meten tot oordelen

Score-cijfer transformatie

De studentscores op de toets moeten omgezet naar een cijfer (vaak op een schaal van 1 tot 10). Dit omzetten gebeurt met behulp van een **score-cijfer transformatieschaal** (zie figuur 2), waarbij de scores in een lineair verband zijn afgezet tegen bijbehorende cijfers.

Het lineair verband wordt gedefinieerd door vastleggen van tenminste twee van de volgende punten.

- A. De score voor het minimale cijfer (bij MC vaak de raadscore = cijfer 1)
- B. De score voor het maximale cijfer (dit kan zijn maximaal haalbare score of de hoogst behaalde scores = cijfer 10);
- C. De cesuurscore (een berekende zak/slaaggrens voor een voldoende (5,5/6,0)).

In het hoger onderwijs is de praktijk veelal dat het cijfer 1 en 10 ongewijzigd blijven, maar dat de cesuurscore wordt aangepast (verlaagd). Transformatie gebeurt dan feitelijk langs twee lijnen met een verschillende richtingscoëfficiënt of te wel een lijn met een 'knik' bij de cesuur. Vanzelfsprekend zal aanpassing van de cesuurscore voor studenten met een score rond de zak-slaaggrens het grootste effect hebben.

Verschillende cesuurmethoden

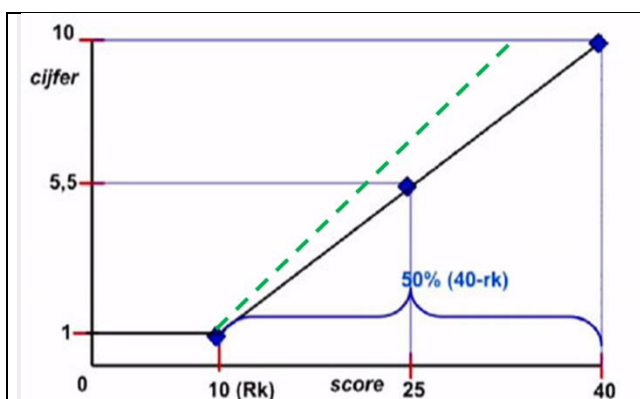
Bij een **absolute cesuurmethode** wordt de zak/slaaggrens al bepaald voordat de studenten de toetsing hebben afgelegd. Het referentiepunt voor de cesuur ligt hierbij dus in de inhoud van het tentamen en de uiteindelijke scores van de studenten hebben daar bij deze methode dan ook geen invloed meer op.

Nadeel: er wordt geen rekening gehouden met de mate van onbetrouwbaarheid van de toets met als gevolg mogelijk onterecht gezakte studenten (te strenge norm).

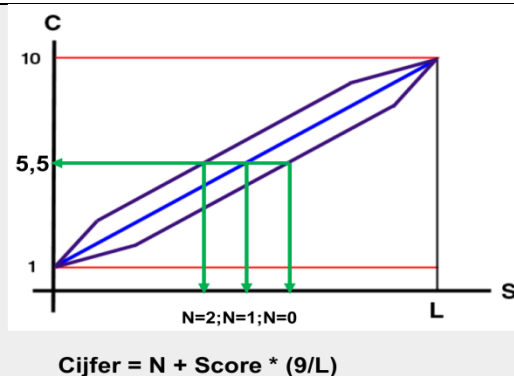
Bij een **relatieve cesuurmethode** ligt het referentiepunt voor de zak/slaaggrens niet in de inhoud van het tentamen maar in de resultaten van een studentenpopulatie.

Nadeel: studenten die zich onvoldoende hebben voorbereid op de toetsing hebben wel invloed op de gemiddelde toetsprestatie met als gevolg mogelijk onterecht geslaagde studenten (te coulante norm).

In de praktijk kom je vaak combinaties van zowel een absolute als relatieve cesuurmethode tegen, die worden ook wel **compromismethoden** genoemd. Hieronder worden de meest gebruikte cesuurmethoden beschreven.



Figuur 2. Links een voorbeeld van een Score-cijfer transformatieschaal van een toets met 40 vierkeuzevragen, een raadkans (Rk) van 10, en een beheersingsgraad van 50% voor een voldoende. De groene stippellijn is een fictief voorbeeld van de methode Cohen-Schotanus, waarbij de hoogst behaalde scores op de toets het referentiepunt zijn voor het hoogste cijfer (in plaats van de maximaal haalbare score). De lijn 'scharniert' als het ware bij de raadkansscore/cijfer 1. De hoogpresteerders op de toets hebben relatief het meeste voordeel bij de aanpassing.



Figuur 3. Rechts een voorbeeld van score-cijfertransformatie zoals gebruikt door het CITO. De N-factor wordt daarbij gebruikt als middel om te corrigeren met 1 cijferpunt boven de N=1 lijn of eronder. Alle studenten gaan er vergelijkbaar op vooruit of op achteruit (m.u.v. de hoogste en laagste prestaties). NB: de scores worden verdeeld over 9 cijferklassen, het cijfer 1 is het startpunt.

Absoluut normeren

1) Vast: Bij een vaste cesuurmethode wordt vooraf aan de toets op basis van de aard van de vragen en een eventuele correctie voor raadkans een cesuur vastgesteld. Het lastige aan deze manier van cesuur bepalen is dat de norm vrij arbitrair is. Er wordt immers niet op basis van inhoudelijke argumenten gekozen voor een bepaalde cesuur.

2) Nedelsky: De Nedelsky methode gaat ook uit van een absolute cesuurmethode, maar hierbij wordt de norm wel beargumenteerd vastgesteld. Nedelsky (1954) stelde voor om een student die een 5,5 zou halen in gedachten te nemen. Vervolgens worden de tentamenvragen door de ogen van deze student bekeken. Daarbij stel je jezelf de vraag: welke antwoordmogelijkheden zou een (hypothetische) 5,5 student kunnen wegstrepen en tussen welke antwoordmogelijkheden zou een 5,5 student twijfelen? Op basis van het aantal antwoordmogelijkheden dat niet wordt weggestreept door een 5,5 student kan de raadkans worden berekend.

Bijvoorbeeld: wanneer van een 5,5 student verwacht wordt dat hij/zij een van de vier antwoordmogelijkheden kan wegstrepen is de raadkans 0,33, aangezien er drie mogelijke antwoorden over blijven.



Door deze raadmans voor iedere vraag te berekenen vanuit het perspectief van de 5,5 student kan een beredeneerde cesuur worden bepaald.

3) Angoff: De cesuurmethode van Angoff (1971) is ook een manier om een beredeneerde absolute cesuur te bepalen. In plaats van het maken van een schatting van hoeveel antwoordmogelijkheden een 5,5 student kan wegstrepen, adviseert Angoff om per vraag te schatten welk percentage van een hypothetische groep 5,5 studenten de vraag goed beantwoordt. Door voor alle vragen de geschatte kansen op te tellen ontstaat een beredeneerde norm.

Relatief normeren

4) Wijnen: De aanname van Wijnen (1971) was dat de moeilijkheidsgraad van een toets in een specifieke context onder specifieke condities alleen afgelezen kan worden aan de scores van studenten. De redenering achter deze cesuurmethode is dan ook dat studenten op basis daarvan beoordeeld moeten worden. In de praktijk betekent dat dat de gemiddelde score van de studenten als uitgangspunt wordt genomen. Deze gemiddelde score min twee standaardmeetfouten geeft een relatief beargumenteerde cesuur. Dit wordt ook wel **grading on the curve** genoemd.

Compromismethode

5) Cohen-Schotanus: Het uitgangspunt bij de methode die is voorgesteld door Cohen-Schotanus en collega's (1996) is in principe een absolute cesuur met een relatief referentiepunt namelijk de groep beste studenten. In plaats van het nemen van de theoretisch hoogst haalbare score kan bijvoorbeeld uitgegaan worden van de daadwerkelijk hoogst behaalde score (met name bij grote groepen studenten).. Ook kan er bijvoorbeeld gekeken worden naar het gemiddelde van de studenten die scoren in het 95^e percentiel en hoger. In de score-cijfer transformatieschaal van figuur zou dan bijvoorbeeld niet 100 als hoogste score worden opgenomen, maar 93 omdat dat de hoogst behaalde score was. Een dergelijke aanpassing is gerechtvaardigd als je mag verwachten dat 'de tien valt'. En dat zal vooral zijn bij eerste afnames waaraan veel studenten meedoen, en waar de kans groot is dat de uitmuntende student in de groep zit. Voor herkansingen en/of kleine groepen is deze methode niet geschikt.

Literatuur voor verdieping

Berkel van, H., Bax A. & Joosten-ten Brinke D. (2014). *Toetsen in het hoger onderwijs*. Houten: Bohn Satfleu van Loghum.

Berkel van, H. Jansen E., & Bax A. (2012). *Studiesucces bevorderen: het kan en is niet moeilijk. Bewezen rendementsverbeteringen in het hoger onderwijs*. Den Haag: Boom Lemma uitgevers

Dousma, T., Horsten, A. & Brants, J. (1997), *Tentamineren*. Groningen: Wolters-Noordhoff

Geraadpleegde literatuur

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Red.), *Educational measurement* (2de editie). Washington, DC: American Psychological Association.

Berkel van, H. Jansen E., & Bax A. (2012). *Studiesucces bevorderen: het kan en is niet moeilijk. Bewezen rendementsverbeteringen in het hoger onderwijs*. Den Haag: Boom Lemma uitgevers

Cohen-Schotanus, J., Van der Vleuten, C. P. M., & Bender, W. (1996). Een betere cesuur bij tentamens. In *Onderzoek van Onderwijs*, 25, 54-55.

Dousma, T., Horsten, A. & Brants, J. (1997). *Tentamineren*. Groningen: Wolters-Noordhoff

Milius, J. (2007). *Schriftelijk tentamineren; Een draaiboek voor docenten in het hoger onderwijs*. Utrecht: IVLOS.

Milius, J., Oost H. & Holleman W. (2002). *Werken aan Academische vorming; Ideeën voor actiefleren in de onderwijspraktijk*. Utrecht: IVLOS.

Nedelsky, L. (1954). Absolute grading standards for objective tests. In *Educational and Psychological Measurement*, 14, 3-19.

Wijnen, W. H. F. W. (1971). *Onder of boven de maat; een methode voor het bepalen van de grens voldoende/onvoldoende bij studietoetsen*. Amsterdam: Swets & Zeitlinger.