

1c Toetslengte, antwoordmogelijkheden en cesuurbepaling

Introductie

De toetsmatrijs (zie handout 1b) geeft inzicht in het aantal vragen nodig om de leerdoelen op een valide en betrouwbare manier te toetsen. In deze handout staat hoe bepaald wordt hoeveel vragen worden opgenomen in een toets (toetslengte). Dit hangt natuurlijk sterk af van de omvang van de te dekken leerstof en ook van de raadkans en het aantal antwoordmogelijkheden. Bovendien is er vaak maar een beperkte periode voor tentaminering beschikbaar. Ook dat bepaalt de toetslengte. Van tevoren wordt ook bepaald wat de cesuur gaat worden. In deze handout wordt stilgestaan bij de volgende vragen:

- Hoeveel vragen komen er in een toets?
- Hoeveel antwoordmogelijkheden zijn er nodig per vraag?
- Hoe wordt de cesuur van de toets bepaald?

Toetslengte en antwoordmogelijkheden

In tabel 1 is een indicatie gegeven van het aantal op te nemen vragen in een toets die een vergelijkbaar meetbereik (30) hebben bij een verschillend aantal alternatieven (twee-, drie- en vierkeuzevragen).

Tabel 1: Relatie tussen aantal alternatieven, toetslengte en meetbereik bij mc-toetsen

Aantal alternatieven(A)	Toetslengte (N)	Raadkansscore (R=N:A)	Meetbereik (N-R)
Vier	40	10	30
Drie	45	15	30
Twee	60	30	30
Twee-, drie-, vier (met raadcorrectie)	30	0	30

Bron: Milius (2007)

Uitleg. Het aantal meerkeuzevragen in de toets is allereerst afhankelijk van de omvang van de leerstof. Verder zal het afhangen of de toets uitsluitend uit meerkeuzevragen bestaat of dat deze vraagvorm wordt gebruikt in combinatie met bijvoorbeeld open vragen. Een toetsmatrijs (of specificatietabel) waarin staat welke inhouden/doelen gedekt worden door welk type vragen kan helpen om te komen tot een toets die in samenstelling de doelen redelijk dekt (zie handout 1b). Daarnaast is het van belang dat de student de vragen in de gestelde tijd kan beantwoorden. Het blijkt dat studenten meer tijd kwijt zijn naarmate het aantal alternatieven toeneemt. Zo is de student gemiddeld 50 seconden kwijt met het beantwoorden van een tweekeuzevraag, 60 seconden met een driekeuzevraag en 75 seconden met een vier- of vijf keuzevraag (Van Berkel & Bax, 2013).

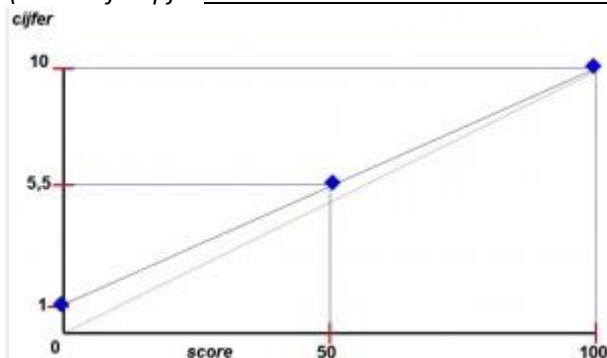
Cesuurbepaling

De cesuur is de zak/slaaggrens bij de beoordeling van een toets. Daarbij stel je jezelf dus de vraag welke score een student tenminste moet behalen om een voldoende te halen of om te slagen. Het aantal punten wat voldoende is voor een voldoende wordt **norm** genoemd.

Studenten kunnen bij een tentamen of toetsing vaak een bepaald aantal punten halen (een score). Vervolgens wordt het aantal behaalde punten of de score omgezet naar een cijfer (vaak op een schaal van 1 tot 10). Dit omzetten doe je met behulp van een **score-cijfer transformatieschaal** (zie figuur). In een score-cijfer transformatieschaal worden de scores die behaald kunnen worden afgezet tegen het cijfer wat daarbij hoort.

De lijn van een transformatieschaal is meestal lineair. Toch zijn er verschillende mogelijkheden om deze lijn te trekken. In andere woorden er zijn verschillende cesuurmethoden die toegepast kunnen worden.

(zie ook filmpje: hoe werkt een absolute cesuurmethode)



Figuur 1. Voorbeeld van een Score-cijfer transformatieschaal

Verskillende 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 omstandigheden waarin de toetsing heeft plaatsgevonden 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 betrouwbaarheid van de toets, de resultaten van een studentenpopulatie uit een eerder cohort of de resultaten van het huidige cohort.

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.

Absoluut

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 raadkans 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

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. (zie handout 5a Toetsanalyse)

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 actief leren 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.