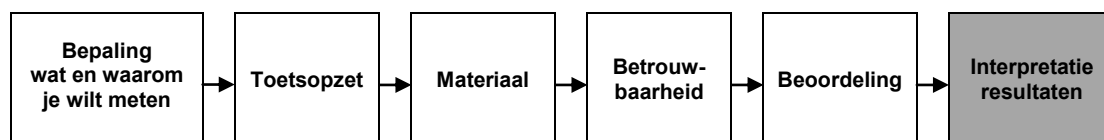


TOETSTIP 10 - JANUARI 2008



TIP 10: CESUURBEPALING

Bij het beoordelen van de taalvaardigheid van studenten komt heel wat kijken: bepalen van wat u wilt meten, zoeken naar materiaal, opstellen van de vragen of de opdracht, een beoordelingsmodel ontwikkelen en het toekennen van scores. Een laatste stap is het interpreteren van de resultaten. In sommige gevallen, zoals bijvoorbeeld bij een eindexamen, is het van belang te bepalen wanneer iemand gezakt of geslaagd is. Dat betekent dat er een zogenaamde cesuur moet worden vastgelegd.

In deze ToetsTip staan we stil bij wat het begrip 'cesuur' nu eigenlijk inhoudt. Vervolgens bespreken we een aantal methoden van cesuurbepaling waarbij we telkens de voor- en nadelen op een rijtje zetten. Daarna lijsten we een aantal tips op met betrekking tot het bepalen van een cesuur in uw eigen toetspraktijk. Traditiegetrouw sluiten we onze ToetsTip af met een opdracht en relevante literatuur over dit onderwerp.

1. Het begrip 'cesuur' ('cut off score', 'standard', 'grenscore', 'drempelscore')

Met de cesuur van een toets bedoelen we die score die de grens vormt tussen een voldoende (= slagen) en een onvoldoende (= zakken). Dat betekent dat studenten met een score gelijk aan of boven de cesuur voldoende hebben gehaald en dus geslaagd zijn voor de toets; studenten met een score onder de cesuur hebben onvoldoende gehaald en zijn dus niet geslaagd op de toets. Als de cesuur bijvoorbeeld op 15 ligt, is een student die 14 heeft gehaald, gezakt en een student die 15 of hoger heeft gehaald, geslaagd.



Het is mogelijk om voor een toets meer dan één cesuur vast te leggen. In dat geval verdeelt u uw groep studenten niet in twee categorieën (gezakt of geslaagd), maar in meerdere categorieën (bv. de studenten die de leerstof heel goed beheersen/net beheersen/net niet beheersen/helemaal niet beheersen).

2. Methoden van cesuurbepaling

Eén van de meest toegepaste manieren om de cesuur van een toets of examen te bepalen is de zogenaamde intuïtieve methode.

Volgens deze methode bepaalt u als docent op basis van uw intuïtie en ervaring waar de grens moet worden gelegd tussen voldoende en onvoldoende.

Vaak wordt een grens van 60% gehanteerd. In dat geval vindt men dat een student iets meer dan de helft van het criterium (bv. de leerstof) moet beheersen en dan komt men op 60% uit. In het geval van het ERK (Europees Referentiekader) wordt als cesuur 75 tot 80% van het criterium gehanteerd.

Het nadeel van deze methode is dat ze niet altijd rekening houdt met de effectieve moeilijkheidsgraad van de toets: de docent kan een bepaalde cesuur vastleggen zonder de moeilijkheidsgraad van de items na te gaan. Een student zou in dat geval een bepaald

percentage moeten halen om te slagen, los van hoe gemakkelijk of hoe moeilijk de toets is. Het spreekt voor zich dat deze methode heel subjectief gekleurd is.

In de literatuur zijn nog heel wat andere methoden beschreven voor het bepalen van de cesuur. Deze methoden kunnen in twee categorieën worden opgesplitst: de item/taak/testgerichte methoden en de prestatie/kandidaatsgerichte methoden. Elk van deze categorieën omvat een aantal concrete methoden. Hieronder bekijken we enkele van deze methoden van naderbij en overlopen we een aantal voor- en nadelen.

2.1 Item/taak/testgerichte methoden

Deze methoden nemen de toets zelf als uitgangspunt om de grens tussen zakken en slagen te bepalen. De meest bekende en gebruikte itemgerichte methode is de modified Angoff methode.

Bij deze methode neemt u een student in gedachten die het criterium (de doelstellingen die u in de toets meet), beheerst en geeft u voor elk toetsitem de kans aan dat deze student het desbetreffende item correct zal beantwoorden. Deze kans is een getal tussen 0 en 1. Nul betekent dat de kans dat de student het item correct oplost onbestaande is en het dus een moeilijk item is. Eén betekent dat de kans dat de student het item correct oplost gelijk is aan 100% en het dus een gemakkelijk item is. Of 0.65 betekent dat de kans dat de student het item correct oplost gelijk is aan 65% is en dus een gemakkelijker item is dan gemiddeld.

De cesuur wordt bepaald door de som te maken van al deze kansen. In het geval van meerdere beoordelaars wordt het gemiddelde of de mediaan van deze grensscores genomen.

Bijvoorbeeld:

Item	Maximale score	Beoordelaar			Gemiddelde
		1	2	3	
1	1	0.33	0.33	0.25	0.30
2	1	1	0.5	1	0.83
3	1	0.5	0.33	0.5	0.44
4	1	0.5	0.5	0.5	0.50
5	1	0.25	0.33	0.33	0.30
Cesuur	5	2.58	1.99	2.58	2.37

Gemiddeld moet een student dus 2,37 van de in totaal 5 punten weten te halen om te slagen op de toets.

Een groot voordeel van de modified Angoff methode is dat ze gemakkelijk uit te voeren is: u hebt er geen speciale software of programma's voor nodig. Bovendien kan deze methode bij elk soort toets gebruikt worden. Zowel bij gesloten vragen (bv. meerkeuzevragen) als bij meer open vragen kan per item de verwachte kans op correcte uitvoering bepaald worden. Op die manier zou u deze methode ook bij schrijf- en spreektoetsen kunnen toepassen.

Toch is het bij deze vaardigheden interessant om een combinatie van test- en kandidaatgerichte methoden te gebruiken en dus om ook concrete schrijf- en spreekprestaties van studenten te gebruiken om de grens tussen slagen en zakken te bepalen (zie verder bij de prestatiegerichte cesuurbepalingsmethoden).

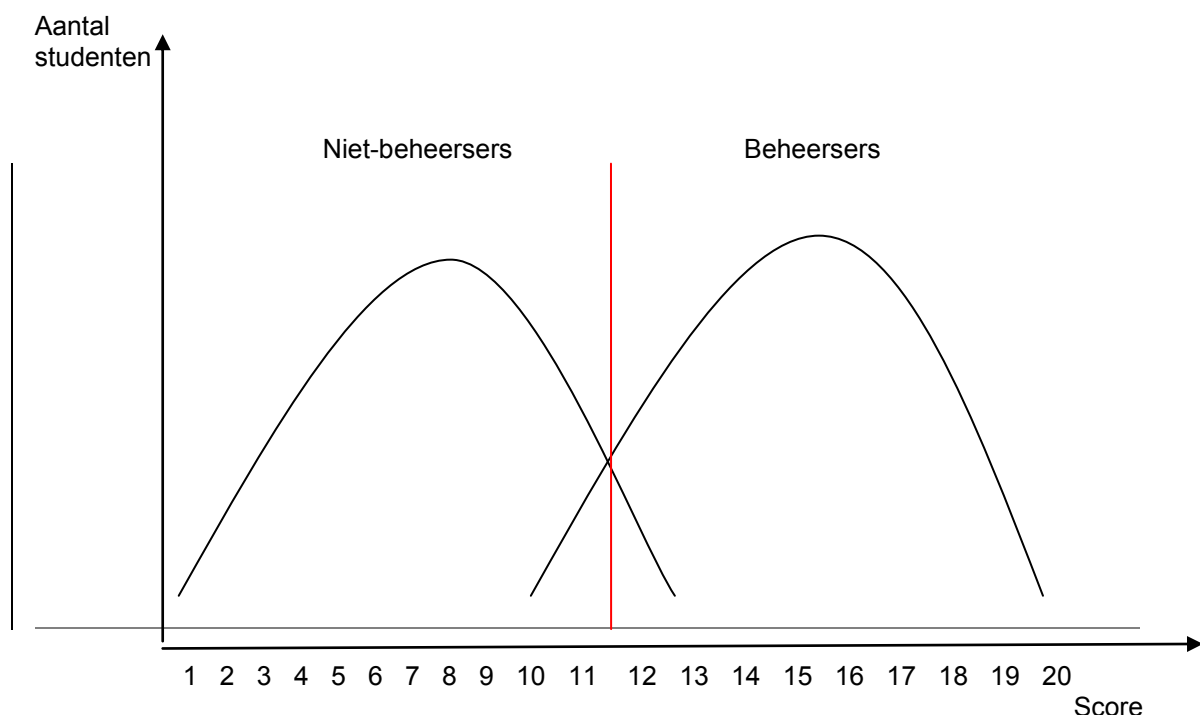
Een belangrijke voorwaarde voor een goede uitvoering van de modified Angoff methode is natuurlijk wel dat u als docent een goed beeld hebt van wat een student die het criterium/de doelstellingen beheerst concreet kan en niet kan. Ook hier speelt subjectiviteit een grote rol in de beoordeling. Het kan daarom aangewezen zijn met meerdere docenten samen de verwachte kansen te bepalen en zo subjectieve factoren in de beoordeling tot een minimum te herleiden.

2.2 Kandidaats/prestatiegerichte methoden

Bij deze methoden staan de prestaties van de studenten op de toets centraal. De cesuur wordt aan de hand van de toetsresultaten van de groep studenten vastgelegd. Binnen de kandidaats/prestatiegerichte methoden bespreken we de methode van de contrasterende groepen en de grens-groep methode, elk met hun voor- en nadelen.

2.2.1 Methode van de contrasterende groepen

Deze methode verloopt in een aantal stappen. Ten eerste verdeelt u uw studenten in twee groepen: de groep 'beheersers' (dit zijn die studenten waarvan u inschat dat ze het criterium/de doelstellingen die de toets meet onder de knie hebben) en de groep 'niet-beheersers' (dit zijn die studenten waarvan u inschat dat ze het criterium/de doelstellingen die de toets meet niet onder de knie hebben). Belangrijk is wel dat u deze inschatting maakt op basis van uw voorgaande (les)ervaringen met de studenten en niet op basis van hun resultaten op de toets. Vervolgens bekijkt u van beide groepen de verdeling van hun scores op de toets.



Ten slotte bepaalt u de cesuur. Dit is die score op de toets die ervoor zorgt dat er zoveel mogelijk studenten uit de groep 'beheersers' slagen op de toets en zo weinig mogelijk studenten uit de groep 'niet-beheersers' slagen op de toets. In de figuur hierboven is dat score 11,5.

De studenten die een score behalen links van de rode verticale lijn (= minder dan 11,5) zijn niet geslaagd op de toets, de studenten die een score behalen op of rechts van de rode verticale lijn zijn wel geslaagd. Het aandeel studenten uit de groep 'beheersers' die niet slagen op de toets is even groot als het aandeel studenten uit de groep 'niet-beheersers' die slagen op de toets.

Een voordeel van de contrast-groep methode is dat ze gemakkelijk te begrijpen en te implementeren is. Ook de berekeningen zijn makkelijk uit te voeren (Berk, 1986). Een ander voordeel van de methode is dat de beoordelaars keuzes moeten maken over studenten die ze kennen. Dit kan de beoordelaars aanspreken (Hambleton, Jaeger & Plake, 2000). Een eventueel nadeel is dat het helemaal niet zo eenvoudig is een studentengroep onder te verdelen in 'beheersers' en 'niet'-beheersers'. Het blijft immers een arbitrair en subjectief onderscheid.

2.2.2 Grens-groep methode

De grens-groep methode is een variant op de methoden van contrasterende groepen. Het verschil is dat u uw studentengroep niet indeelt in een groep 'beheersers' en 'niet-beheersers', maar wel dat u binnen uw studentengroep die studenten selecteert waarvan u vindt dat ze net wel/net niet het criterium/de doelstellingen beheersen. Dit gebeurt opnieuw los van de effectieve scores op de toets.

De mediaan (dit is de middelste score of het gemiddelde van de twee middelste scores) is dan de cesuur van de toets.

Het belangrijkste nadeel van deze methode is dat het selecteren van de studenten die net wel/net niet het criterium/de doelstellingen beheersen, een subjectieve zaak blijft. Overleg met collega-docenten kan de invloed van deze subjectieve factoren enigszins minimaliseren.

3. Tips en bedenkingen bij het vastleggen van een cesuur

- Meer methodes: Dé cesuurbepalingsmethode bestaat niet, elke methode heeft voor- en nadelen. Door verschillende methoden met elkaar te combineren kunnen de nadelen tot een minimum herleid worden.
- Consequenties: Hou bij de keuze van de methode rekening met de aard en zwaarte van de beslissingen die op basis van de cesuur genomen zullen worden en met de beschikbare middelen (tijd en geld).
- Meer oordelen: Hou er rekening mee dat de genomen beslissingen in zekere mate altijd arbitrair en subjectief blijven. Door met collega's of andere ervaringsdeskundigen te overleggen, kan een deel van deze subjectiviteit weggewerkt worden.

4. Interessante literatuur

Berk, R.A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56 (1), 137-172.

Berkel, H. van (1999). *Zicht op toetsen. Toetsconstructie in het hoger onderwijs*. Assen: Van Gorcum

Cizek G.J. (1996). Standard-setting guidelines. *Educational Measurement: Issues and Practice*, 15 (1), 13-21.

Cizek, G.J. (2001). *Setting performance standards: Concepts, methods and perspectives*. London:Lawrence Erlbaum Associates, Inc.

Hambleton, R.K., Jaeger, R.M., Plake, B.S. & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, 24 (4), 355-366.

Plake, B.S. (1998). Setting performance standards for professional licensure and certification. *Applied Measurement in Education*, 11 (1), 65-80.

<http://www.education.ualberta.ca/educ/psych/crame/files/RickerCSSE2003.pdf>

OPDRACHT

- Evalueer uw eigen cesuurbepalingsmethodes. De volgende vragen kunnen hierbij als richtlijn dienen:

- Welke van bovenstaande beschreven methode(s) hebt u in uw loopbaan als docent Nederlands als Vreemde Taal al gehanteerd?
- Wat vond u zelf de voor- en nadelen van deze methode(s)?
- Welke methode(s) geniet(en) uw voorkeur en waarom?

- Hanteer bij een volgende toetsafname een cesuurbepalingsmethode die u nog nooit hebt gebruikt. Vergelijk deze met de methode(s) die u anders gebruikt. Wat zijn de voor- en nadelen van elke methode?

U kunt uw ervaringen met deze opdracht naar ons doorsturen voor feedback. Wanneer u vragen hebt over dit onderwerp, kunt u deze natuurlijk ook altijd aan ons stellen. Onze contactgegevens staan onder aan deze ToetsTip.

TOETSTIP 10 CESUURBEPALING JANUARI 2008

In de volgende ToetsTip gaan we dieper in op de culturele bepaaldheid van toetsen

Vragen, opmerkingen en suggesties

Opmerkingen over ToetsTip 10:

Suggesties voor volgende ToetsTips:

Inhoudelijke opmerkingen over evaluatie in het algemeen:

Praktische bemerkingen over de ToetsTip:

Voor vragen & reacties één adres:

het CNaVT-secretariaat

Certificaat Nederlands als Vreemde Taal
Blijde-Inkomststraat 7
B-3000 Leuven
België
0032 (0) 16 53 55 16

cnavt@arts.kuleuven.be