

EEN BETERE CESUUR BIJ TENTAMENS: DE BESTE STUDENTEN ALS REFERENTIEPUNT

J. Cohen-Schotanus¹, C.P.M. van der Vleuten² en W. Bender¹

Samenvatting

De manier waarop de grens tussen slagen en zakken wordt vastgesteld (de cesuur) verschilt per vak en/of opleiding. Veel gebruikte cesuren zijn de absolute cesuur (meestal 60% gecorrigeerd voor raden) en de relatieve cesuur (de gemiddelde groepsprestatie minus de standaarddeviatie). Bij de medische faculteit in Groningen is men aanhanger van de absolute cesuur, in Maastricht gaat de voorkeur uit naar relatief meten. Beide cesuren hebben hun voor- en nadelen. Bij de absolute cesuur wordt bijvoorbeeld geen rekening gehouden met toevalligheden waaraan een tentamen bloot staat, met als mogelijk gevolg ten onrechte gezakte studenten. Een nadeel van de relatieve cesuur is dat studenten die zich niet voldoende hebben ingezet wel hun invloed hebben op de gemiddelde groepsprestatie met als mogelijk gevolg ten onrechte geslaagde studenten. Om uit deze impasse te komen wordt een nieuwe cesuur voorgesteld die aan beide nadelen tegemoet komt: een absolute cesuur met een relatief referentiepunt, te weten de beste studenten.

Inleiding

Bij de discussies over de kwaliteit en de doelmatigheid van het hoger onderwijs gaat het vooral over het studierendement en de doorstroom. Immers, met doelmatigheid en kwaliteit wordt vooral bedoeld een hoger studierendement en een snellere doorstroom. Manieren om de kwaliteit van het onderwijs te verbeteren worden gesuggereerd in het studeerbaarheidsrapport van de commissie Wijnen.¹ Opvallend is dat in dat rapport relatief weinig aandacht wordt besteed aan de kwaliteit van de toetsing in het hoger onderwijs, terwijl de tentamenuitslagen toch direct van invloed zijn op het studierendement. Het lijkt wel alsof het vanzelf spreekt dat het met de kwaliteit van de toetsing wel goed zit. Dat is echter niet het geval en de aantallen studenten die ten onrechte voor tentamens zakken zijn niet gering.²

Een tentamen is kwalitatief goed als voldaan is aan voorwaarden als representativiteit, relevantie, objectiviteit, eenduidigheid, etcetera.³ Verder moet de beoordelingsprocedure voor studenten inzichtelijk zijn. Hoe minder tentamens aan genoemde voorwaarden voldoen, hoe moeilijker tentamens voor studenten zijn. Voorbeelden van wat er bij tentamens mis kan gaan zijn er te over:

- toetsvragen zijn onduidelijk geformuleerd;
- toetsvragen zijn niet representatief voor de stof;
- de toetsvragen gaan de opgegeven stof te buiten;
- het ene jaar worden moeilijker vragen gesteld dan het andere jaar;
- etcetera.

Daarnaast is het mogelijk dat onduidelijkheden en/of onzorgvuldigheden in de onderwijssituatie zoals een slechte planning van het toetsmoment, slecht onderwijs, een te laat uitgegeven klapper of een uitverkocht boek van invloed zijn op het tentamenresultaat. In alle gevallen kan men zich afvragen of het terecht is dat (een aantal) studenten hier het slachtoffer van worden. Met de hierboven gesignaleerde problemen kan rekening gehouden worden door de methode waarmee de grens tussen een voldoende en een onvoldoende prestatie (cesuur) wordt vastgesteld, van een amendement te voorzien.

¹Rijksuniversiteit Groningen, ²Rijksuniversiteit Limburg

Relatieve versus absolute cesuur

Op het eerste gezicht lijkt het hanteren van een relatieve cesuur te corrigeren voor onzorgvuldigheden in de toetsing en/of het bijbehorende onderwijs. Wijnen beargumenteert in zijn proefschrift dat het groepsgemiddelde de beste voorspeller is van wat de deelnemende groep, in de gegeven onderwijssituatie, bij dit tentamen kan presteren. Hij suggereert daarom als cesuur te nemen het groepsgemiddelde minus twee maal de standaardmeetfout.⁴ Wijnen gaat er daarbij vanuit dat alle aan het tentamen deelnemende studenten het onderwijs hebben gevolgd en het tentamen naar behoren hebben voorbereid. De methode van Wijnen corrigeert dus voor ruis aan de onderwijs- en de tentamenkant. Maar in de praktijk kan daarbij sprake zijn van enige overcorrectie. Meestal is het collegebezoek lang niet 100% en zijn er altijd studenten die wel aan het tentamen meedoen maar zich niet hebben voorbereid.² Het groepsgemiddelde ligt dan eigenlijk te laag. In een dergelijke situatie is de 'regel van Wijnen' ongeschikt omdat er dan studenten zullen zijn die ten onrechte slagen. Bij de medische faculteit in Groningen werd daarom, na een aantal jaren gebruik te hebben gemaakt van de 'regel van Wijnen', massaal overgestapt op een absolute cesuur (60% kennis gecorrigeerd voor raden).

In Groningen had gebruik van de absolute cesuur tot gevolg dat het gevraagde kennisniveau een stabiele factor was, maar dat, bij hetzelfde vak, het ene jaar veel meer studenten zakten dan het andere jaar, terwijl de inspanning van de groep studenten ongewijzigd was. In Maastricht blijkt het percentage gezakte studenten vrij stabiel te zijn, maar is de variatie in het gevraagde kennisniveau groot.⁵ De Visitatiecommissie Geneeskunde en Gezondheidswetenschappen concludeerde dan ook in Groningen dat "de kwaliteitsbewaking van de toetsen te kort schiet: deze dienen regelmatig bewaakt te worden". In Maastricht concludeert de commissie dat "alles bij elkaar genomen het beoordelingssysteem mild en kwetsbaar is".⁶ In beide faculteiten wordt gezocht naar een alternatief voor de gehanteerde cesuur.

Alternatieve cesuur: de beste student als referentiepunt

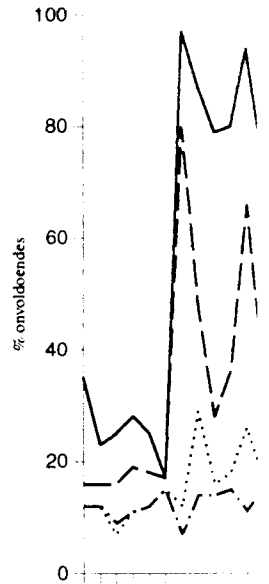
Een aantal Groningse vakken heeft als tussenoplossing de door Hofstee gesuggereerde cesuur toegepast. Vrij vertaald komt deze cesuur er op neer dat binnen bepaalde grenzen geschoven wordt met de cesuur, afhankelijk van het percentage gezakten.⁷ Deze cesuurberekening blijkt in de praktijk lastig te zijn voor veel docenten. Daarbij komt dat veel docenten gevoelsmatig een voorkeur hebben voor een absolute cesuur.

Een mogelijkheid om uit de bovenbeschreven impasse te raken is het volgende. Het principe is een absolute cesuur (bijvoorbeeld 60% kennis) met een relatief referentiepunt. In plaats van de maximale score als referentiepunt te nemen, wordt de prestatie van de beste studenten als referentie beschouwd. Bijvoorbeeld als de beste studenten gemiddeld 90% van de vragen goed hebben, dan komt de cesuur te liggen bij 60% van de 90%, gecorrigeerd voor raadkans. De aanname is dat de beste studenten het onderwijs volgen zoals dat de bedoeling is en zich de opgegeven stof naar behoren eigen maken. Anders gezegd, de onzorgvuldigheid die er ook aan de studentkant kan bestaan, doet zich bij de beste studenten niet voor. Door de beste studenten als referentiepunt te nemen, wordt rekening gehouden met de toevallige moeilijkheid en de ruis van de toets, zonder dat de 'niet serieuze studenten' daar enige invloed op uitoefenen.

De cesuren in de praktijk

Zowel in Maastricht⁸ als Groningen is van reeds afgenomen tentamens berekend wat de uitslag zou zijn geweest bij een andere cesuur. Bij beide faculteiten werden vergelijkbare uitkomsten gevonden. Voor de verdere redenering worden de Groningse resultaten gebruikt. Nagegaan is wat het percentage onvoldoendes zou zijn geweest bij een absolute cesuur (60% kennis gecorrigeerd

voor raden, de relatieve cesuur en bij een absolute cesuur. De absolute cesuur met uit te gaan van de hoogste score van de beste student presteert (uitschieter naar boven) is voor de score waarop Een analyse van 60 Groningse de vier cesuren op (figu



Figuur 1 Percentages

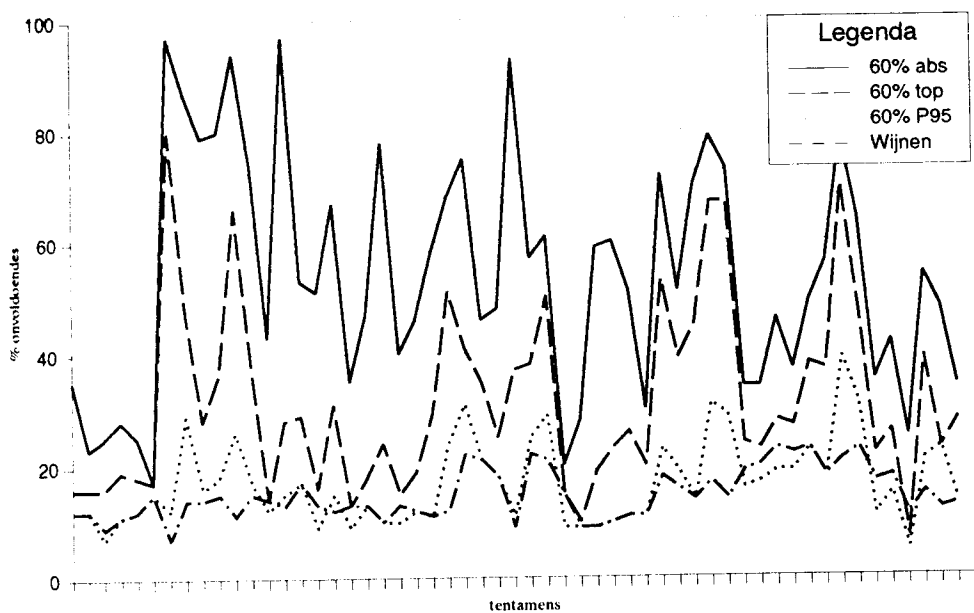
Duidelijk is dat de fluctuaties in de absolute cesuur en de absolute cesuur uitslagen is bij de absolute cesuur minder. De variatie in de gevonden uitslagen gaat uit naar een absolute cesuur voorkeur heeft te maken met fluctuaties in de uitslagen.

Wanneer de uitslagen van de percentage onvoldoendes duidelijker te zien dat de

voor raden, de relatieve cesuur van Wijnen (gemiddelde minus twee maal de standaard meetfout) en bij een absolute cesuur met een relatief referentiepunt.

De absolute cesuur met een relatief referentiepunt is op twee manieren berekend. Ten eerste door uit te gaan van de hoogste prestatie. De cesuur wordt dan berekend door 60% te eisen van wat de beste student presteert (gecorrigeerd voor raden). Omdat ook bij de beste student het toeval een uitschieter naar boven kan veroorzaken, is daarnaast een ander referentiepunt genomen. Gekozen is voor de score waarop het 95-ste percentiepunt valt.

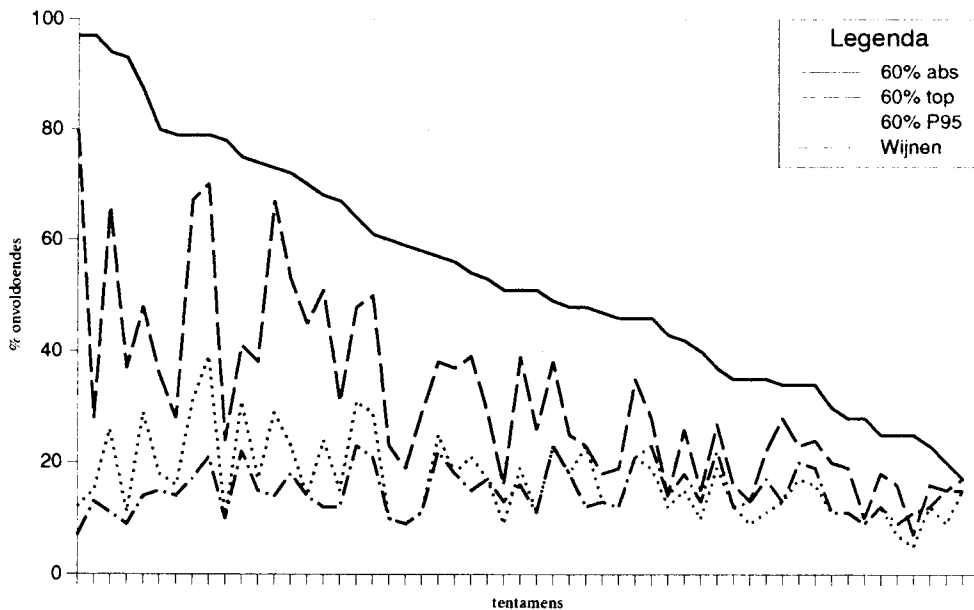
Een analyse van 60 Groningse toetsen, in volgorde van afname, leverde zeer diverse uitslagen bij de vier cesuren op (figuur 1).



Figuur 1 Percentages onvoldoendes bij vier verschillende cesuren bij zestig tentamens

Duidelijk is dat de fluctuaties in tentamenuitslagen groot zijn, vooral bij de 'gewone' absolute cesuur en de absolute cesuur met de best presterende student als referentiepunt. De variatie in uitslagen is bij de absolute cesuur met het 95-ste percentiepunt als uitgangspunt aanzienlijk minder. De variatie is bij de uitslagen onder de regel Wijnen het kleinst. Op basis van de gevonden uitslagen gaat de Groningse voorkeur, bij afwijzing van de regel van Wijnen, voorlopig uit naar een absolute cesuur met het 95-ste percentiepunt als referentiepunt. Die Groningse voorkeur heeft te maken met de opvatting dat er geen enkele reden is om aan te nemen dat de fluctuaties in de uitslagen bij de absolute cesuur, reële niveauverschillen tussen groepen weerspiegelen.

Wanneer de uitslagen van de absolute cesuur van 60% worden gesorteerd op volgorde van het percentage onvoldoendes, ontstaat een overzichtelijker grafiek (figuur 2). In deze grafiek is nog duidelijker te zien dat de relatieve cesuur van Wijnen de minste fluctuaties in uitslagen laat zien.



Figuur 2 Percentage onvoldoendes bij vier verschillende cesuren, gesorteerd op de 60% absolute cesuur, bij zestig tentamens

Totaal hebben ongeveer 10.000 studenten meegedaan aan de in deze analyse opgenomen tentamens. Het gemiddeld percentage gezakte studenten zou bij de beschreven cesuren als volgt zijn geweest (zie tabel 1).

Tabel 1 Gemiddeld percentage onvoldoendes bij vier verschillende cesuren.

cesuur	% onvoldoendes N = 10.000
60% absoluut	54
60% top student	36
60% van P95	17
Wijnen	15

Gemiddeld bedraagt het verschil in percentages onvoldoendes tussen de 'gewone' absolute cesuur en de regel Wijnen bijna 40%. In werkelijkheid is het verschil iets minder groot omdat sommige docenten bij zeer slechte tentamenresultaten de cesuur bijstellen.

Beschouwing

Het is duidelijk dat cesuur - heel wat n ook niet verwonder een jaar korter is d van de Maastrichts examens scoren de collega's van de and Het is de vraag of tentamens, als gevo als Maastricht is de studentgedrag te ve onzorgvuldigheid a percentiel als refere Men kan zich afvr: opbrengst geweest studietijd en dus 10 als referentiepunt z miljoen gulden heb bedrag feitelijk wat uitslagen de cesuur absolute cesuur, vo De gegevens die hi is, gezien de studie studierichtingen bet het de hoogste tijd zorgvuldigheid van is het aan te beve referentiepunt.

Literatuur

- 1 Wijnen WHF studeerbaarheid Ministerie va
- 2 Cohen-Schott: kennis en stu Rijksuniversi
- 3 De Groot AI ren. Den Ha
- 4 Wijnen WHI voldoende/or gen, 1971.
- 5 Vleuten CPN

Beschouwing

Het is duidelijk dat medische studenten in Maastricht - als gevolg van de gehanteerde relatieve cesuur - heel wat minder herkansingstentamens afleggen dan studenten in Groningen. Het is dan ook niet verwonderlijk dat de verblijfsduur van geneeskunde studenten in Maastricht gemiddeld een jaar korter is dan die van hun collega's in Groningen. Dit betekent geenszins dat het niveau van de Maastrichtse studenten lager ligt. Zowel op schriftelijke toetsen als bij vaardigheids-examens scoren de Maastrichtenaren op hetzelfde niveau als de Groningse studenten en hun collega's van de andere zusterfaculteiten.^{9 10}

Het is de vraag of het nodig is af te stappen van een relatieve cesuur, want de ruis rond de tentamens, als gevolg van onzorgvuldigheden aan de onderwijstkant, is groot. Zowel in Groningen als Maastricht is de grote fluctuatie in tentamenuitslagen bij een absolute cesuur niet met wisselend studentgedrag te verklaren.^{2 5} Als men niet relatief wil meten, maar toch wil corrigeren voor de onzorgvuldigheid aan de onderwijstkant dan lijkt op dit moment de absolute cesuur met het 95-ste percentiel als referentiepunt het beste alternatief.

Men kan zich afvragen wat, als deze cesuur vijf jaar eerder zou zijn bedacht en toegepast, de opbrengst geweest zou zijn. Het overdoen van een vak kost een student minimaal een maand studietijd en dus 1000 gulden. Het hanteren van de absolute cesuur met het 95-ste percentiepunt als referentiepunt zou voor de betrokken Groningse studenten tezamen een besparing van een paar miljoen gulden hebben opgeleverd. Volledigheidshalve moet daar aan toegevoegd worden dat het bedrag feitelijk wat lager uit gevallen zou zijn omdat sommige docenten bij desastreuze tentamenuitslagen de cesuur hebben aangepast. Verder zal elke overstap van een relatieve naar een meer absolute cesuur, volgens deze rekenwijze, eerder verlies dan winst opleveren.

De gegevens die hier aan de orde komen hebben betrekking op de studierichting geneeskunde. Er is, gezien de studiedoorstroom in Nederland, weinig reden aan te nemen dat de situatie bij andere studierichtingen beter is. Ook daar zal de ruis om de tentamens heen groot zijn. Onzes inziens is het de hoogste tijd dat er, als het om de kwaliteit van onderwijs gaat, gekeken wordt naar de zorgvuldigheid van de procedures rondom toetsing en tentamens. Zolang die situatie niet ideaal is, is het aan te bevelen om het gebruik van een absolute cesuur te corrigeren met een relatief referentiepunt.

Literatuur

- 1 Wijnen WHFW, Wolthagen HAP, Bie D de, et al. Te doen of niet te doen? Advies over de studeerbaarheid van onderwijsprogramma's in het Hoger Onderwijs. Leiderdorp: Graficon, Ministerie van O&W, 1992.
- 2 Cohen-Schotanus J. Effecten van curriculumveranderingen, studiewaardering, studeergedrag, kennis en studiedoorstroom in een veranderend medisch curriculum [dissertatie]. Groningen: Rijksuniversiteit Groningen, 1994.
- 3 De Groot AD, Naerssen RF van (redactie). Studietoetsen construeren, afnemen en analyseren. Den Haag: Mouton, 1969.
- 4 Wijnen WHFW. Onder of boven de maat. Een methode voor het bepalen van de grens voldoende/onvoldoende bij studietoetsen [dissertatie]. Groningen: Rijksuniversiteit Groningen, 1971.
- 5 Vleuten CPM van der, Hoogenboom RJI, Wijnen WHFW. Relatieve normen in examen-

- beslissingen: een steen des aanstoots. Voordracht Gezond Onderwijscongres (GOC IV), 1994.
- 6 VSNU-Visitatiecommissie Geneeskunde en Gezondheidswetenschappen. Onderwijsvisitatie Geneeskunde en Gezondheidswetenschappen. Utrecht: VSNU, 1992.
 - 7 Hofstee WKB. Een gebufferde oplossing voor het bepalen van de grens tussen voldoende en onvoldoende. Universiteit en Hogeschool 1982; 28:21-8.
 - 8 Muijtjens AMM, Hoogenboom RJI, Vleuten CPM van der. Absolute normering bij voortgangs- en bloктоetsen. EPG-publ.nr. 95-35. Maastricht: Rijksuniversiteit Limburg, 1995.
 - 9 Bender W, Cohen-Schotanus J, Imbos Tj, Versfelt WA, Verwijnen GM. Medische kennis bij studenten uit verschillende faculteiten: van hetzelfde laken een pak? Ned Tijdschr Geneeskd 1984; 128:917-21.
 - 10 Louw A de, Verryn Stuart AA, Keuning FM, Scherpbier AJJA, Pols J. Hoe goed ben ik voorbereid? Vaardigheidsonderwijs in Maastricht en Groningen getoetst aan Raamplan 1994 en de praktijk. In: Pols J, Cate ThJ ten, Houtkoop E, Pollemans MC, Smal JA. Gezond Onderwijs-4. Houten/Zaventem: Bohn Stafleu Van Loghum, 1995.

Trefwoorden

cesuurproblemen
kwaliteit toetsing
studievoortgang

Correspondentieadres

Dr. J. Cohen-Schotanus
Bureau Onderzoek van Onderwijs Geneeskunde
Faculteit der Geneeskunde
Antonius Deusinglaan 1
9713 AV Groningen

Samenvatting
*In september .
curriculum in
vergeleken me
programma pe
De studentpar
feitelijk bestee
geprogramme*

Inleiding

In september
van start ge
traditionele o
studenten var
gaat hierbij a
vragenlijsten
tijdsbesteding
175 studente
uitkomsten v

Het traditio

Kenmerken

De kenmerk
meeste onde
Ochtends w
keer was er
vooral op d
trimester w
wijs georga
voeding en
was, was ze

Student- en

Het oordee
negatief: m
voor de lat
oordeel ove
het kijkje a
stukje erva
en een goe

Faculteit d