

(Artikel) Compensatorisch toetsen

Thea van Lankveld, Silvester Draaijer ¹

Onderwijscentrum Vrije Universiteit

Oktober 2010

Inleiding

Een examenregeling waarbij studenten verplicht zijn alle onderwijsonderdelen met een voldoende af te sluiten, komt veel voor in het Nederlandse hoger onderwijs. Een dergelijke regeling wordt een conjunctieve regeling genoemd.

Naast conjunctieve regelingen bestaan ook compensatorische regelingen. Bij een compensatorische regeling kunnen studenten de goede resultaten bij het ene onderdeel gebruiken als compensatie bij een onvoldoende voor een ander onderdeel.

In dit hoofdstuk zijn op basis van de literatuur de voor- en nadelen van compensatorische examenregelingen op een rijtje gezet.

Voordelen van compensatie (de theorie)

In de literatuur over examenregelingen, wordt als voornaamste voordeel van compensatie genoemd dat deze leidt tot een snellere studievoortgang (o.a. Jansen, 1996; Rekveld & Starren, 1994; Visser-Wijnveen, 2010). Hiervoor worden drie argumenten genoemd.

Het eerste argument komt voort uit de psychometrische literatuur (o.a. De Gruijter, 2008). Daarin wordt compensatie vooral verdedigd als middel om foute examenbeslissingen te verminderen. Dat zit zo: in elk tentamen zit, vanwege onduidelijke vraagstelling en/of fouten in de beoordeling, wel een bepaalde onnauwkeurigheid. Daardoor kunnen studenten ten onrechte zakken (of ten onrechte slagen, dat kan ook). Voor de groep studenten die onnodig moet herkansen treedt bij gevolg studievertraging op. En naarmate het aantal tentamens toeneemt, zal de kans op deze (onnodige) studievertraging groter worden. Compensatie kan een instrument zijn om het aantal van dergelijke foute beslissingen te verminderen.

Een tweede voordeel dat wordt genoemd, is dat een compensatieregeling, zeker wanneer die samen gaat met weinig herkansingen (bij sommige opleidingen zijn er zelfs geen herkansingen), de studenten stimuleert om hard te studeren en niet uit te stellen. De regeling stimuleert namelijk om voor een hoog cijfer te gaan, want extra inspanningen worden in het systeem beloond. Het behalen van een hoog resultaat voor de eerste toets vergroot namelijk de kans van slagen voor de tweede toets. Bij een conjunctieve regeling daarentegen is een studiestrategie die leidt tot een zes voor een tentamen het meest effectief, want de kans van slagen is het hoogst als men de studie-inspanningen goed verdeelt over de verschillende onderdelen.

¹

Met dank aan Wouter Mulders en Wim van Os voor hun commentaar op een eerdere versie van dit stuk.

Een derde, meer pragmatisch voordeel van compensatie is dat het zal leiden tot minder deelname aan herkansingen (Rekvelde & Starren, 1994). Dat betekent minder werk voor studenten en minder werk voor docenten. In deze context wordt er door deze auteurs ook op gewezen dat het leereffect van herkansingen zelden hoog is; studenten blijken de leerstof nauwelijks beter te beheersen bij een herkansing. Dus leidt een herkansing voornamelijk tot tijdverlies van zowel docenten als studenten, en nauwelijks tot leerwinst.

Samenvattend, de voordelen van compensatie bestaan er vooral uit dat een dergelijke regeling leidt tot een snellere doorstroming en dus beter studiesucces.

De praktijk van compensatie

De gedachte dat een compensatieregeling leidt tot betere studievoortgang wordt in empirisch onderzoek bevestigd. Drie studies in het Nederlandse hoger onderwijs laten dit zien, één bij de Geneeskunde in Groningen (Cohen-Schotanus, 1995), één bij Economie in Rotterdam (Arnold & Van den Brink, 2009) en één grootschalig onderzoek bij meerdere opleidingen in Nederland (Van den Berg & Hofman, 2005; Jansen, 1996).

Bij de faculteit Geneeskunde aan de Universiteit Groningen mochten studenten voor hun doctoraal-examen per studiejaar één vijf laten staan, mits deze vijf gecompenseerd werd door een zeven op een ander vak (Cohen-Schotanus, 1995). Dit is een vorm van gedeeltelijke compensatie (zie ook verderop). In theorie kunnen studenten hier voor hun doctoraal afstuderen met drie vijven op de lijst (per studiejaar één). Door de invoering van de compensatieregeling verminderde de studieduur met ong. drie maanden.

In de studie van Cohen-Schotanus (1995) werd meer in detail gerapporteerd hoe vaak studenten dan gebruik maakten van de compensatieregeling, en of het leidde tot ander studeergedrag. 61% Van de doctoraalstudenten bleek gebruik te maken van de regeling; zij lieten gemiddeld 1,7 vijven staan. En de meeste studenten compenseerden deze behaalde vijf ruimschoots; in hun cijferlijst kwamen meestal meerdere zevens, achten en negens voor. Uit deze studie bleek dat er geen standaard patroon was waarbij een 'moeilijk vak' werd gecompenseerd door een 'gemakkelijk vak'. Alle tentamens werden wel eens gecompenseerd. En het ene jaar was het weer een ander vak dan het andere jaar. Wel wordt het ene vak vaker gecompenseerd dan het andere, maar Cohen-Schotanus wijdt dit meer aan de toetsvormgeving dan aan strategisch gedrag van studenten: het vak waarbij het vaakst een vijf is blijven staan was in die jaren aan het experimenteren met nieuwe tentamenvormen, waarbij het voor de studenten vaak ondoorzichtig was wat nou de bedoeling was van het vak. Ook andere vakken waarbij iets mis ging bij het tentamen (veel vragen buiten de stof, vragen die niet pasten bij de nieuwe uitgave van het boek) werden vaker gecompenseerd. Cohen-Schotanus concludeert dan ook dat er waarschijnlijk sprake is van compensatie van 'toevallige' vijven.

Bij de Erasmus School of Economics (Arnold & Van den Brink, 2009) werd in 2007 een nieuw curriculum ingevoerd, waarvan een gedeeltelijke compensatieregeling onderdeel was. Studenten mochten in het eerste en tweede bachelorjaar binnen een cluster van verwante vakken van in totaal 20 EC, één vijf compenseren, mits het gewogen gemiddelde in het cluster voldoende is. Daarnaast werd het aantal herkansingen ingeperkt: studenten mochten voor maximaal drie van de tien vakken herkansen (per bachelorjaar). Sinds de invoering van deze regeling is het aantal positieve BSA-adviezen (studenten die 60 EC halen in één jaar) gestegen van 28% naar 43%. Het aantal negatieve BSA's (studenten die minder dan 40 EC halen) is gelijk gebleven. Het aantal studenten dat de maximale compensatiemogelijkheid benutte (drie onvoldoendes) was zeer klein; 2,6%. 57,1% Van de studenten

maakte geen gebruik van compensatie, 40,4% compenseerde één of twee onvoldoendes. Verder bleek dat compensatie niet leidde tot slechtere prestaties in het tweede jaar. Met andere woorden: deze compensatieregeling lijkt vooral positief uit te pakken voor de studenten die over het geheel genomen goed presteren, zonder dat het ten koste gaat van hun studiestatistiek later in de studie. Voor de studenten die minder dan 40EC halen lijkt de regeling geen effect te hebben.

Recenter bleek ook bij andere Nederlandse opleidingen waar compensatieregelingen gehanteerd werden, het studiesucces hoger te zijn dan bij opleidingen met een conjunctief regime (Van den Berg & Hofman, 2005; Jansen, 1996).

Nadelen van volledige compensatie

Er zijn echter ook bezwaren in te brengen tegen een compensatorische regeling.

Het meest gehoorde bezwaar is dat een laag cijfer op een belangrijk vak gecompenseerd kan worden door een hoog cijfer op een minder belangrijk vak. Op die manier kan een student het examen halen zonder dat hij/zij van een belangrijk vak wellicht voldoende weet. De aanname bij een conjunctieve regeling is eigenlijk dat ieder vak even belangrijk en onmisbaar is en dat alleen 'voor alle vakken voldoende' het bewijs levert dat 'het eindniveau' bereikt is. Tegen deze aanname kan ingebracht worden dat het eindniveau niet bepaald wordt door 'de som van de vakprestaties', maar dat er (als het goed is) in het handelen en denken van studenten geïntegreerde nieuwe kennis ontstaat – hopelijk in het verlengde van de algemene opleidingsdoelen.

Daarnaast wordt de vrees gehoord dat studenten vakken zullen verwaarlozen, omdat ze met goede cijfers voor andere vakken toch wel zullen slagen (bijvoorbeeld bij het laatste vak: "ik ben toch al geslaagd") (Van Os, 1995).

Beide bezwaren kunnen worden ondervangen met gedeeltelijke compensatie (De Gruijter, 2008).

Gedeeltelijke compensatie

Om de hierboven genoemde nadelen van volledige compensatie te ondervangen, zou gedeeltelijke compensatie kunnen worden overwogen.

Mogelijkheden van gedeeltelijke compensatie zijn:

- Het bepalen van een ondergrens. Bijvoorbeeld dat alleen vijven mogen worden gecompenseerd. Dit is een manier om toch een minimale kwaliteit van het diploma (en dus van de afgestudeerden) te borgen.
- Het beperken van het aantal onvoldoendes dat mag worden gecompenseerd (bijvoorbeeld één onvoldoende per jaar, zoals bij Geneeskunde in Groningen het geval was). Een ondergrens aanbrengen in het aantal onvoldoendes wordt gewenst geacht omdat het belangrijk is te voorkomen dat studenten met overwegend onvoldoendes toch kunnen slagen. In zo'n geval zou met evenveel recht vraagtekens gezet kunnen worden bij de 'terechtheid' van de zevens als bij de terechtheid van de vijven. Anders gezegd: het moet gaan om een min of meer toevallige onvoldoende binnen een veld van voldoende.
- Compensatie alleen toestaan binnen een cluster van vakken (bijvoorbeeld een onvoldoende op een M&T-vak mag alleen worden gecompenseerd door een voldoende op andere M&T-vakken). Ook dit is een manier om te garanderen dat een bepaalde minimale kwaliteit wordt gegaran-

deerd, maar deze optie zou een hindernis *kunnen* vormen voor studenten die voor het overige goed presteren.

- De compensatieregeling alleen laten gelden voor een bepaalde periode, bijvoorbeeld voor het eerste jaar dat een student voor een bepaalde studie is ingeschreven. Een dergelijke regeling zal studenten stimuleren zich in te zetten voor de studie (en dus studiesucces).

Zak/slaaggrens bepaling

Een interessant aspect van Toetsen en Rendement betreft het aantal studenten dat op een toets zakt op basis van de cesuur (zak/slaaggrens).

Vanuit de optiek van redelijkheid voor studenten kan worden gesteld dat het ernstiger is dat studenten onterecht zakken dan dat zij onterecht slagen. In het kader van het verhogen van het rendement van het onderwijs geldt die stelregel zeker. Het kost immers extra tijd en geld om studenten die eigenlijk al geslaagd zijn, nog een keer onderwijs te laten volgen en/of te examineren.

Om zinvolle uitspraken te doen over wat verdedigbare zak/slaaggrenzen zijn, kan gesteld worden dat er in het algemeen drie methoden zijn te onderscheiden. Géén een van deze methoden is de 'juiste' methode. Op basis van redelijke argumenten zal de zak/slaaggrens uitgewerkt moeten worden voor elke cursus.

Absoluut	Bij de absolute methode bepaalt u vooraf de voldoende/onvoldoende grens. Die grens wordt ingegeven door het uitgangspunt dat gecontroleerd moet worden of studenten voldoen aan de eisen die kunnen worden afgeleid uit de doelstellingen van het onderwijs.
Relatief	De relatieve methode gaat uit van het idee dat de toets afgestemd moet zijn op studenten die gerechtigd zijn tot het volgen van onderwijs. Op basis van dat principe moet de meerderheid van hen in staat geacht worden om te slagen voor een toets. Omdat vooraf niet bekend is waartoe studenten bij het <i>gegeven</i> onderwijs en de <i>gegeven</i> toets in staat zijn, kan de norm <i>niet</i> vooraf worden vastgesteld. Het resultaat op de toets moet worden afgewacht om de grens te kunnen vaststellen.
Compromis	De compromismethode probeert de principiële verschillen tussen absoluut en relatief cesuurstellen te overbruggen en heeft in de onderwijspraktijk de voorkeur. Bij de compromismethode wordt in het algemeen uitgegaan van een absolute norm en wordt aangegeven onder welke omstandigheden van deze absolute norm zal worden afgeweken. Een van de compromismethoden staat hieronder beschreven.

Open vragen	Gesloten vragen
<ol style="list-style-type: none"> 1. Begin met het bepalen van de grens voldoende/onvoldoende. Het is niet ongebruikelijk die te leggen bij 55% of 60% van de maximaal te behalen score. 2. Indien u vindt dat er teveel studenten zakken voor uw toets, zijn (na overleg met de examencommissie) de volgende aanpassingen in 	<ol style="list-style-type: none"> 1. Begin met het bepalen van de grens voldoende/onvoldoende. Het is gebruikelijk deze bij 60% van de hoogste score te leggen met inachtneming van de kansscore. <ul style="list-style-type: none"> • De kansscore is het aantal geldige vragen in de toets gedeeld door het aantal alternatieven per vraag. Bijvoorbeeld de

<p>de procedure mogelijk.</p> <ol style="list-style-type: none"> a. In plaats van 60% stelt u de grens op 55% van de maximale score en/of b. U neemt niet de maximale score als uitgangspunt, maar het gemiddeld van de vijf hoogste scores. <p>3. U heeft nu de ruwe score die een 0 oplevert en de ruwe score die een 10 oplevert: de overige cijfers bepaalt u door de ruwe scores op te delen in 10 gelijke eenheden met daaraan gekoppeld een cijfer.</p>	<p>kansscore met 105 juist/onjuistvragen is $105/2=52,5$.</p> <ul style="list-style-type: none"> • Als bij de genoemde toets de hoogste score 96 punten is (dus geen enkele student heeft alle vragen correct), komt de voldoende/onvoldoende grens te liggen bij 60% van de score tussen 52,5 en $96 = 78,6$. <ol style="list-style-type: none"> 2. Studenten met de hoogste score (96) krijgen een 10, zij met de kansscore een 0. Met behulp van een lineaire regressie zijn de overige cijfers te berekenen. 3. Indien u vindt dat er teveel studenten zakken voor uw toets: zijn na overleg met de examencommissie de volgende aanpassingen in de procedure mogelijk: <ol style="list-style-type: none"> a. In plaats van 60% stelt u de grens op 55% van de maximale score en/of b. U neemt niet de maximale score als uitgangspunt, maar het gemiddelde van de vijf hoogste scores.
--	--

Misclassificaties

Gegeven de bepaalde zak/slaaggrens, kan de vraag gesteld worden hoeveel studenten onterecht zakken of slagen, gegeven de kwaliteit van de toets. Dit aspect betreft het aantal misclassificaties. Met de kwaliteit van een toets wordt in dit verband bedoeld de *psychometrische betrouwbaarheid* van de toets (dat wil zeggen de berekende betrouwbaarheid van een toets: bijvoorbeeld Cronbach alfa). Zou een student bij een herhaalde afname van de toets dezelfde score behalen? Hoeveel ruis en error zit er eigenlijk in de meting die gedaan is?

Hoe hoger deze betrouwbaarheid van een toets, hoe lager het aantal misclassificaties. De betrouwbaarheid van een toets wordt vooral bepaald door de kwaliteit en het aantal vragen in een toets, de consistentie van beoordeling (vooral van belang bij open vragen) en de spreiding van de kennis onder de studentenpopulatie.

Er zijn meer of minder complexe methoden om het aantal misclassificaties te schatten. Hieronder volgt een uiteenzetting van de meest eenvoudige methode (Sanders, in voorbereiding).

Ten eerste dient men te beseffen dat de score die een student op een toets behaalt, bestaat uit zijn ware score (de werkelijke score van de student) plus error (ruis veroorzaakt door bijvoorbeeld onduidelijke vragen, omstandigheden, toeval etc). Dit is één van de belangrijkste uitgangspunten in de toetstheorie. Deze ruis kan uitgedrukt worden in de standaard meetfout (SE). Met behulp van de standaard meetfout is het mogelijk om de grenzen te bepalen waarbinnen zich een percentage van de score zich met een bepaalde zekerheid bevindt. Er zijn een aantal manieren om deze meetfout te bepalen, maar bij een 'normale mc-toets' (gemiddelde score ongeveer 60% van de maximale score), kan deze worden geschat door de wortel te trekken van de maximumscore op een toets en deze te vermenigvuldigen met 0,45.

Als we bijvoorbeeld een toets nemen met een maximum score van 100 punten, dan is de standaard meetfout de wortel uit $100 \cdot 0,45 = 4,5$ punten. Stel nu dat een student een score behaalt (een geobserveerde score) van 62 punten. Stel dat de cesuur 67 punten bedraagt. Hoe 'waar' is dan de uitspraak dat de student terecht gezakt is? Niet heel terecht moeten we concluderen want in dit geval ligt de ware score van een student met 90% zekerheid in het interval tussen $62 - 4,5 \cdot 1,65 = 54,6$ punten en $62 + 4,5 \cdot 1,65 = 69,4$ punten. Met andere woorden, de student had net zo goed kunnen slagen.

Gegeven dit interval voor de ware score van een student, kan ook bepaald worden welke score met 90% betrouwbaarheid als ware score hoger is dan de zak/slaaggrens $67 + 2 \cdot 4,5 \cdot 1,65 = 81,9$ punten. Andersom geredeneerd kan bepaald worden bij welke score met 90% betrouwbaarheid kan worden aangegeven dat iemand terecht is gezakt: $67 - 2 \cdot 4,5 \cdot 1,65 = 52,1$ punten.

Zakslaaggrens toets	67 punten
90% betrouwbaar terecht gezakt	52,1 punten
90% betrouwbaar terecht geslaagd	81,9 punten

Het is duidelijk dat de scores verkregen op toetsen – zeker rond de zak/slaaggrens – met de nodige voorzichtigheid moeten worden behandeld.

Op basis van de bovenstaande redenering kan ook bepaald worden welk percentage van de studenten in principe onterecht zakt cq. slaagt. Het voert te ver om dat hier uit te werken, maar onderstaande tabel 7.2 is overgenomen uit Dousma et al. (1997). De tabel geeft het percentage misclassificaties bij een gegeven zakpercentage (afwijzingspercentage).

Tabel 7.3 bevat de gegevens voor een toets met bijvoorbeeld 200 studenten, een betrouwbaarheid van 0,70 en 30% onvoldoendes.

Tabel 7.2 Percentages niet-consistente beslissingen als functie van afwijzingspercentage en (paralleltoets)betrouwbaarheid.

Afwijzings- percentage	Betrouwbaarheid											
	0,0	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	1,00
5	10	8	7	7	7	6	6	5	4	4	3	0
10	18	14	13	12	11	11	10	9	8	6	4	0
15	26	18	18	17	16	14	13	12	10	8	6	0
20	32	23	21	20	19	17	16	14	12	10	7	0
25	38	26	25	23	22	20	18	16	14	11	8	0
30	42	29	27	25	24	22	20	18	15	12	9	0
35	46	31	29	27	25	23	21	19	16	13	9	0
40	48	32	30	29	27	24	22	20	17	14	10	0
45	50	33	31	29	29	25	23	20	18	14	10	0
50	50	33	31	30	27	25	23	20	18	14	10	0

Tabel 7.3

	geslaagd	gezakt	totaal
geslaagd	118	22	140
gezakt	22	38	60
totaal	140	60	200

Tot slot

Of men nu tot compensatie besluit of niet, het is altijd mogelijk om aan de hand van de studievoortgangsgegevens in de database door te rekenen wat een andere examenregeling dan de huidige, zou betekenen voor de studievoortgang van de huidige populatie studenten. Op die manier is het mogelijk zicht te krijgen op het effect van de verschillende regelingen. Hetzelfde geldt voor bijvoorbeeld het verhogen van de betrouwbaarheid van toetsen door het toevoegen van toetsvragen en het dientengevolge verlagen van het aantal misclassificaties.

Voor welke optie een opleiding ook kiest, het is altijd aan te raden vooraf deze doorrekening te maken.

Reacties op dit artikel zijn welkom op <http://blog.han.nl/hantd2010/compensatorisch-toetsen>

Literatuur

- Arnold, I.J.M., & van den Brink, W. (2009). De invloed van compensatie op studie-uitval en doorstroom. TH&MA, 3, 11-15.
- Van den Berg, M.N., & Hofman, W.H.A. (2005). Student success in university education: A multi-measurement study of the impact of student and faculty factors on study progress. Higher Education, 50, 413-446.
- Cohen-Schotanus, J. (1995). De praktijk van compensatie. Onderzoek van onderwijs, 24, 60-62.
- Dousma, T., Horsten, A., & Brants, J. (1997). Tentamineren. Hoger Onderwijsreeks (derde druk., Vol. 3). Wolters-Noordhoff.
- De Gruijter, D.N.M. (2008). Al dan geen compensatie in de propedeuse. Retrieved June 4, 2010 from Leiden University, ICLON website: http://wordpress.iclon.nl/HO/wp-content/uploads/2009/09/Al_dan_geen_compensatie_propedeuse_DatodeGruijter_2008.pdf

- Jansen, E.P.W.A. (1996). Curriculumorganisatie en studievoortgang: een onderzoek onder zes studierichtingen aan de Rijksuniversiteit Groningen. Dissertatie Universiteit Groningen.
- Os, W. van (1995). Compensatieregelingen. Onderwijsadviesbureau, Vrije Universiteit Amsterdam.
- Rekveld, & Starren, (1994). Een examenregeling zonder compensatie in het Nederlandse Hoger Onderwijs? Een vergelijking tussen compensatie en conjunctie. Tijdschrift voor Hoger Onderwijs, 12(4), 20-29.
- Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. Practical Assessment, Research & Evaluation, 7(14). Retrieved from <http://pareonline.net/getvn.asp?v=7&n=14>
- Sanders, P. (2010) Toetsen op School. Cito. In voorbereiding.
- Visser-Wijnveen, G. (2010). Inrichting toetsing. Retrieved June 4, 2010, from: <http://www.studiesuccesho.nl>