

Incompatibiliteit van toetsing in het hoger onderwijs

Dr. H. Starren is werkzaam bij de Vakgroep Psychologie van de Rijksuniversiteit Groningen.

(j.c.m.g.starren@ppsw.rug.nl)

Toetsen dienen te stimuleren tot maximale kwaliteit. En kwaliteit betekent in dit verband dat zoveel mogelijk studenten zoveel mogelijk leren dat (later) van belang is. De technische term voor die eis is compatibiliteit. In de universitaire praktijk wemelt het van de aanwijzingen voor incompatibiliteit: het lukt zelden om toetsen zo te maken en te gebruiken dat ze studenten uitnodigen om het onderste uit de kan te halen. Waarom dat zo is illustreer ik met recent onderzoek naar de wijze waarop propedeuse-studenten Psychologie RuG zich op tentamens plegen voor te bereiden. Daaruit blijkt dat een fors aantal studenten dat voor tentamens slaagt zich volstrekt onvoldoende inspant. Dit kwaliteitsprobleem dient opleidingen een zorg te zijn. Er is alle reden voor Examencommissies om van meer absolute methoden van cesuurbepaling gebruik te gaan maken en om de kennis-eisen om te kunnen slagen voor tentamens fors aan te scherpen.

VOORAF

Het is zo langzamerhand een eufemisme om te zeggen dat er wat schort aan de wijze waarop we aan de universiteit toetsen maken en gebruiken, speciaal in de onderbouw van grote faculteiten. Met onze hang naar meettechnische perfectie en efficiëntie lopen we het risico te weinig oog te tonen voor de stimulerende aspecten van toetsing. Waar dat toe kan leiden valt gemakkelijk te illustreren aan de wijze waarop wij multiple-choice tentamens maken en gebruiken in de propedeusefase van grote faculteiten: studenten worden onvoldoende gestimuleerd om de leerstof kritisch en creatief te verwerken (Vermunt, 1992; Anderson, 1997; Hofstee, 1999a). Elders voerde ik daarom een pleidooi om daar waar dat mogelijk is de beoordelaar weer 'in beeld te brengen', zoals bij mondelinge tentamens en het beoordelen van scripties het geval is. De kans op strategisch gedrag van studenten is bij dergelijke toetsvormen miniem vanwege het doorbre-

ken van de anonimiteit van de beoordelingssituatie. In zulke gevallen blijkt een grondige voorbereiding de meest doelmatige manier om een positief oordeel te oogsten (Starren, 1998).

De veronderstelling dat je met toetsing daadwerkelijk iets kunt veranderen aan het studiegedrag is herhaaldelijk in onderzoek geverifieerd (Mc Kenzie, 1973; Marton et al., 1976; Boud, 1995; Gibbs, 1999). Tentamens werpen hun schaduw vooruit: studenten blijken bij het voorbereiden van tentamens te anticiperen op de eisen waaraan ze moeten voldoen om van de docent een positieve beoordeling te krijgen. Dat betekent dan ook dat het normstellende karakter van toetsen sterk richtinggevend is voor de wijze van voorbereiden en daarmee voor de uitkomst van die leeractiviteiten van studenten. Tentamens fungeren dus niet alleen als afhankelijke variabelen maar zijn in elk onderwijssysteem tevens werkzaam als belangrijke onafhankelijke variabelen.

Welnu, een rationeel bestuurd onderwijsorganisatie hanteert de toetsing dan ook bewust. Dat betekent dat het beoordelingssysteem zo is ingericht dat studenten studeren zoals dat zou moeten. Die eis heet compatibiliteit, een term uit de micro-economie, vergelijkbaar met feedforward uit de systeemtheorie.

AANWIJZINGEN VOOR INCOMPATIBILITEIT?

Aanwijzingen voor incompatibiliteit van toetsing zijn er genoeg. De huidige beleidspraktijk laat zien dat het autoriteiten in grote faculteiten maar moeilijk lukt om zich te verplaatsen in het gezichtspunt van degenen die onderhevig zijn aan beoordelingsmaatregelen. Die pakken dan ook vaak verkeerd uit: de beoordeling is niet verenigbaar met het doel, de maatregelen leiden tot ongewenste neveneffecten (Hofstee, 1999b). Een bekend voorbeeld is de beoordeling van vaardigheden. Sommige opleidingen kennen vaardigheidsproeven die meetellen in het examensysteem; andere opleidingen volstaan met papieren simulaties, ook als het om vaardigheden gaat die het meest aangegeven lijken op directe observatie (het houden van een interview of de afname van een verhoor). Meettechnische motieven geven wellicht de doorslag. De gevolgen laten zich raden: studenten blijken zich in hun voorbereiding eerder te richten op de theorie over vaardigheden dan op de training van de vaardigheden zelf. Daar waar praktische vaardigheidstoetsen geïntroduceerd worden treden belangrijke verschuivingen in de studieactiviteiten van studenten op: studenten blijken zich minder te richten op het bestuderen van de theorie en meer op het oefenen van de onderwezen vaardigheden (Smit, 1995; Anderson, 1997; Gibbs, 1999).

Een ander voorbeeld is het gebruik van tussentijdse toetsen. Het doel ervan is nobel: men wil studenten vroegtijdig confronteren met resultaten zodat ze in staat zijn hun werkwijze aan te passen. Maar de student weet wel beter. Die neemt alleen deel aan tussentijdse toetsen als die vrijstellend karakter hebben voor de afsluitende toets over de tentamenstof. Of iets motiveert, wordt nu eenmaal niet bepaald door de vraag of iemand dat leuk en interessant vindt maar of het een bijdrage levert aan het halen van het tentamen.

INCOMPATIBEL GEBRUIK VAN DE SLAAG/ZAK-GRENS

Een ander voorbeeld waaraan zich het gezichtspunt van compatibiliteit gemakkelijk laat illustreren is de cesuurbepaling. Zo bestaat er in brede kring, ook bij de betere studenten, ongenoegen over het lage kennispercentage waarmee men voor tentamens kan slagen. Bij veel tentamens ligt de feitelijke cesuur zodanig dat amper meer dan 40% reële kennis (score gecorrigeerd voor giskans) geëist wordt. Dat lage kennispercentage zou iedere examencommissie een doorn in het oog moeten zijn.

Tegelijkertijd zijn de zakpercentages vaak hoog. Zakpercentages boven de 60%, speciaal in de propedeuse, zijn bijvoorbeeld bij mijn eigen vakgroep, de vakgroep Psychologie in Groningen, niet ongewoon. Ook dat is een stand van zaken die iedere examencommissie onrustig zou behoren te maken.

Men kan het laten bij de constatering dat toetsen, vooral in de onderbouw van massa-opleidingen, nu eenmaal beroerd gemaakt worden. Of, nog erger, vervallen tot cynisme, zoals een medische faculteit enkele jaren geleden, die haar propedeuserendement opkrikte door tentamencijfers met een halve punt te verhogen. Men zou echter ook de kwaliteit van tentamenvragen in de beschouwing kunnen betrekken. En tegelijkertijd kunnen nagaan of een verhoging van de cesuur niet eerder te rechtvaardigen is dan een verlaging van de kenniseisen, de gebruikelijke reflex in gevallen van hoge zakpercentages. Want er is alle reden om weerstand te bieden aan de druk die van dergelijke zakpercentages pleegt uit te gaan: studenten klagen niet zelden over toetsen die weinig inspirerend zijn en die amper uitdagen tot grote dadendrang.

De Groningse vakgroep Psychologie zocht de oplossing in een heroverweging van de cesuurbepaling waarbij zowel de moeilijkheidsgraad van de tentamenvragen als de hoogte van de kenniseisen in de beschouwing betrokken zijn, met hopelijk positieve gevolgen voor het onderwijs en voor de inspanning van studenten. Voor verhoging van de eisen om voor tentamens te kunnen slagen voelde de opleiding wel, maar men was toch wat huiverig voor de consequenties van dat voorstel (men was namelijk bang voor absurd hoge percentages onvoldoenden). Daarom leek het raadzaam om naar de afzonderlijke tentamenvragen te kijken en alleen die vragen te betrekken bij het bepalen van de uitslag die niet te moeilijk blijken voor de serieuze tentaminandi (het kan niet zinnig de bedoeling van de examinerende geweest zijn om items op te nemen die door de serieuze tentaminandus onvoldoende worden gemaakt).

Alvorens in te gaan op onze ervaringen met de nieuwe aanpak, is het nuttig om iets meer te zeggen over de manier waarop de norm bij Psychologie RuG wordt vastgesteld. De norm komt tot stand via een model dat elke docent bekend in de oren zal klinken. We zoeken naar een compromis tussen wat we enerzijds een maximaal toelaatbaar percentage onvoldoenden achten en, anderzijds, een minimum percentage kennis dat we nog aanvaardbaar vinden. De parameters van dit compromismodel zijn voorwerp van beleid en worden door de Examencommissie dan ook regelmatig heroverwogen.

De invulling van de huidige variant luidt: 'stel de cesuur zodanig in dat $(K)ennispercentage + 1/3 \times (Z)akpercentage = 70$, onder de randconditie van minimaal 50% kennis en maximaal 60% zakpercentage'. Ook deze variant, zo'n twee jaar oud, geeft weinig reden om tevreden te zijn. Het resultaat is dat nogal wat zakpercentages in de propedeuse rond de 50% liggen, alhoewel het ook regelmatig voorkomt dat de cesuur niet

meer automatisch bepaald kan worden (Z groter dan 60%). Niettemin is ook in die gevallen de minimum kenniseis van $K=50\%$ gehandhaafd omdat er bij nader inzien niet aan de deugdelijkheid van de betreffende tentamens hoefde te worden getwijfeld. Opvallend is overigens dat de hogere kenniseis gemiddeld zeker niet leidde tot het gevreesde rendementsverlies. De zakpercentages bleven praktisch onveranderd, namelijk beroerd hoog. Dat kan op twee manieren geduid worden. Ofwel, met die hogere eisen valt het bij nader inzien mee, bijvoorbeeld omdat de itemselectie-methode er toe leidt dat items verwijderd worden die gaan over moeilijke delen van de stof, en niet, zoals de bedoeling is, slaan op die items die bij nader inzien niet deugen, bijvoorbeeld vanwege een verkeerde formulering of een fout in de sleutel (en daarvoor zijn inderdaad aanwijzingen gevonden). Ofwel, de studenten pasten hun studie-inspanning aan bij die hogere eisen en in dat geval rijst de vraag waarom de opleiding hier niet nog meer gebruik van zou maken.

Er is dus nog steeds sprake van incompatibel gebruik van de slaag/zak-grens: de zakpercentages zijn immers onveranderd hoog en er is twijfel over de kenniseisen. En om de onaanvaardbare lage eisen was het Psychologie RuG toch allemaal begonnen, niet zozeer om de zakpercentages? Kennelijk is het ook met de 'nieuwe' eisen niet gelukt om een meer adequate voorbereiding uit te lokken.

De vraag rijst dan ook of het niet tijd wordt voor forsere maatregelen. Waarom bijvoorbeeld geen kenniseis van circa 70% stellen om te kunnen slagen (overeenkomend met $K=70\%$, hetgeen bij een tweekeuzetentamen correspondeert met minstens 85% goede antwoorden)? Wat is daar eigenlijk op tegen? Zo'n eis strookt bijvoorbeeld heel goed met de gedachte dat opleidingen (zoals Psychologie RuG) alleen een boodschap zouden moeten hebben aan de norm-student, iemand dus die qua inzet de volle werkweek aan de studie besteedt en qua intelligentie gemiddeld begaafd is (Zelfstudie Psychologie RuG, 2000).

VOORBEREIDINGSWIJZE OP PROPEDEUSETENTAMENS

Normverhoging biedt het risico op forse stijging van zakpercentages, zeker voor de korte termijn. Doorvoering is niettemin verdedigbaar als aannemelijk gemaakt kan worden dat slechte resultaten niet zozeer het gevolg zijn van te hoge eisen van examinatoren of te weinig aandacht in het onderwijs voor moeilijke onderwerpen maar eerder van onvoldoende voorbereiding van studenten.

Omdat het zakpercentage in het compromisemodel van Psychologie RuG als buffer fungeert voor al te grote fluctuaties in de moeilijkheidsgraad van tentamens (via de Z -max parameter), leidt bovenstaande overweging tot de vraag of het zakpercentage wel een adequate schatting oplevert van die moeilijkheidsgraad. Waarom zou je rekening houden met een maximaal percentage onvoldoenden als blijkt dat een groot aantal zakkers zich volstrekt onvoldoende heeft ingespannen? En waarom zou je het minimaal vereiste kennisniveau niet kunnen opkrikken als blijkt dat zelfs geslaagden achterblijven bij de normatieve opvatting over studeren?

Voor een tweetal propedeusevakken uit de opleiding Psychologie RuG is onlangs nagegaan of er inderdaad zulke goede gronden zijn voor een stevige normverhoging.

Onderzoeksvragen

Met het oog op de beslissing over wat acceptabele kennis- en zakpercentages zijn, is het antwoord op de volgende twee vragen van belang:

Is het hoge zakpercentage het gevolg van onvoldoende tentamenvoorbereiding?

Zit er voldoende speling in de tentamenvoorbereiding van geslaagden om verhoging van de kennis-eis te rechtvaardigen?

Als blijkt dat hoge zakpercentages grotendeels te wijten zijn aan onvoldoende tentamenvoorbereiding, dan kan het maximaal toelaatbare zakpercentage (Z-max in het compromisemodel) hoger gesteld worden. En als blijkt dat dat zelfs de geslaagden qua voorbereidingstijd en de aard van de voorbereiding serieus achterblijven bij wat men zou mogen verwachten, dan kan dit een goede reden zijn om het minimaal vereiste kennispercentage (K-min in het compromisemodel) te verhogen.

Bij tentamenvoorbereiding gaat het ons op de eerste plaats om de tijd die besteed is aan het voorbereiden van tentamens. Enerzijds, omdat het puntenaantal voor tentamens indicatief is voor de hoeveelheid voorbereidingstijd die men van een student mag verwachten (één studiepoint komt overeen met een bruto studielast van veertig uur). Anderzijds, omdat de hoeveelheid bestede tijd een belangrijke predictor is van studieresultaten, zeker op individueel niveau (Schouwenburg, 1996). Op de tweede plaats gaat het ons ook om de aard van de voorbereiding. De student wordt, bijvoorbeeld, geacht de literatuur grondig te bestuderen en niet te volstaan met bestudering van andermans uittreksels. Hofstee (1998a) bestempelt het voorbereiden op tentamens door middel van niet zelf gemaakte uittreksels zelfs als bedrog, omdat deze praktijk een zuivere vaststelling van de mate waarin de student zich in de stof verdiept heeft in de weg staat.

Werkwijze

Om de onderzoeksvragen te beantwoorden is achtereenvolgens bij twee propedeusevakken navraag gedaan naar de tentamenvoorbereiding: Overzicht Sociale Psychologie (in 1998/99) en Statistiek 1 (in 1999/00). Beide vakken tezamen zijn qua moeilijkheidsgraad en inhoud redelijk representatief voor de propedeuse van psychologieopleidingen. Afzonderlijk zijn ze elk drie studiepunten waard, zodat de normatieve studielast voor beide 120 uur bedraagt.

Na afloop van het tentamen (eerste kans) is schriftelijk een korte lijst voorgelegd met vragen naar de bestede voorbereidingstijd en de aard van de voorbereiding. Bij Sociale Psychologie vulden 189 van de 346 deelnemers (54%) de lijst bruikbaar in, bij Statistiek 1: 259 van de 402 deelnemers (64%). Op aselectheid is gecontroleerd door de tentamencijfers van de onderzoeksgroepen te vergelijken met die van de overige tentamendeelnemers. Noch de gemiddelden, noch de standaarddeviaties verschilden substantieel: bij Sociale Psychologie zijn die gemiddelden en standaarddeviaties resp. 5.8 (s.d. 1.5) voor de onderzoeksgroep en 5.7 (s.d. 1.5) voor de overigen; bij Statistiek 1 is dat: 4.6 (s.d. 1.9) voor de onderzoeksgroep en 4.3 (s.d. 1.9) voor de overigen.

Zowel bij Sociale Psychologie als Statistiek 1 is gevraagd naar de bestede tijd en de wijze van voorbereiden. Om een zo precies mogelijke schatting te krijgen van de totaal geïnvesteerde tijd is afzonderlijk gevraagd naar de zelfstudietijd en naar het (werk)collegebezoek.

De vraag naar de zelfstudietijd is ontleend aan Cohen-Schotanus (1994), omdat de gekozen formulering een nauwkeurige schatting van het aantal studie-uren oplevert. De formulering luidt: 'Kun je schatten hoeveel tijd je in de voorbereiding van vak X gestoken hebt? Probeer de schatting zo nauwkeurig mogelijk te maken. Een goede manier is eerst te schatten hoeveel uur je gemiddeld per dag voor dit tentamen gestudeerd hebt. Vermenigvuldig dat vervolgens met het aantal dagen dat je met ... bezig bent geweest. Laat de (werk)collegetijd buiten beschouwing'.

Daarnaast is er bij Statistiek 1 nog gevraagd naar de hoeveelheid bestudeerde tentamenstof, zowel halverwege als aan het eind van de cursus. Een dergelijke vraag zegt overigens ook al iets over de kwaliteit van het leren omdat mag worden aangenomen dat gelijkmatiger spreiding van de tentamenstof over de cursusduur er toe leidt dat de stof beter kan worden opgenomen. Aan de wijze van voorbereiden is vervolgens voor elk vak een specifieke meerkeuzevraag gewijd. Op basis van de antwoorden zijn de studenten in groepen ingedeeld, waarbij onderscheid is gemaakt tussen 'eigenlijke' en 'oneigenlijke' manieren van studeren. Voorbeelden van het eerste zijn: 'ik heb het boek een of meer keer grondig doorgenomen', 'Ik bereidde me voor door sommen te maken'. Oneigenlijk noemden we manieren van voorbereiden als: 'Ik volstond met het leren van oude tentamens' en 'ik bereidde me voor door alleen andermans uittreksel te leren'.

De vragenlijsten zijn direct met het tentamen uitgedeeld. Daarbij was een korte instructie gevoegd met het verzoek ook het studentnummer in te vullen zodat we in staat waren de antwoorden te koppelen aan het tentamencijfer. De respons was redelijk en qua tentamencijfers nauwelijks verschillend van de niet-onderzochten (zie eerder).

Resultaten

Tentamenvoorbereiding is op de eerste plaats uitgedrukt in de tijd die studenten hebben gestoken in het studeren. Voor beide vakken staat een normatieve studielast van 120 uur. De eerste vraag is de vraag of het hoge percentage zakkers het gevolg is van onvoldoende voorbereiding voor het tentamen. Van de onderzochte groepen blijkt voor Sociale Psychologie 46% gezakt; voor Statistiek 1 is dat 70%. Voor Sociale Psychologie heeft deze groep gemiddeld 33 uren (hoorcollege- en zelfstudie-uren) besteed aan de voorbereiding, dus 87 minder dan de normatieve 120. Bij het gevonden gemiddelde ligt het steekproefgemiddelde met 95% zekerheid tussen 28 en 38 uur. De gezakten bij Statistiek 1 hebben gemiddeld 50 uren (werkcollege- en zelfstudie-uren) aan de voorbereiding op het tentamen besteed, en dat is 70 uur minder dan de normatieve 120 uur. Het 95%-betrouwbaarheidsinterval voor het steekproef-gemiddelde ligt tussen 45 en 55 uur. Een voorlopig antwoord op de tweede onderzoeksvraag 'Zit er voldoende speling in de tentamenvoorbereiding van geslaagden om verhoging van de kenniseis te rechtvaardigen' wordt gegeven door het gemiddeld aantal uren voorbereiding van de studenten die geslaagd zijn. Dit aantal is voor Sociale Psychologie 48 uur, voor Statistiek 1 is dat 62 uur. Ook deze aantallen zijn beduidend lager (72 resp. 58 uren) dan de norm van 120. In de tabellen 1a en 1b wordt een overzicht gegeven van de resultaten.

Tabel 1a: Voorbereidingstijd gezakten en geslaagden Sociale Psychologie: gemiddelden, standaarddeviaties, 95%-betrouwbaarheidsintervallen

	Gemiddelde	Standaarddeviatie	Betrouwbaarheidsinterval
Gezakten (N=86; 46%)	33	23	28 - 38
Geslaagden (N=103; 54%)	48	23	43 - 52

Tabel 1b: Voorbereidingstijd gezakten en geslaagden Statistiek 1: gemiddelden, standaarddeviaties, 95%-betrouwbaarheidsintervallen.

	Gemiddelde	Standaarddeviatie	Betrouwbaarheidsinterval
Gezakten (N=182; 70%)	50	35	45 - 55
Geslaagden (N=77; 30%)	62	39	53 - 71

Aangezien de gezakten en geslaagden bij lange na niet het normatieve urenaantal van 120 hebben besteed, is het maar zeer de vraag of ze de tentamenstof überhaupt volledig hebben doorgenomen. Alleen voor Statistiek 1 zijn daarover gegevens verzameld. En, inderdaad, zelfs de geslaagden blijken gemiddeld genomen aan het eind van de cursus niet aan de gehele stof toegekomen (71%, tegenover 54% van de stof bij de gezakten). Halverwege de cursus heeft de geslaagde 35%, de gezakte 27% van de tentamenstof bekeken. Behalve over de tijd die gemoed is met de voorbereiding kan er ook iets gezegd worden over de manier waarop gestudeerd is. Van de student wordt verwacht dat hij het boek grondig bestudeert, sommen oefent, en zo mogelijk een eigen uittreksel maakt. We noemden dat eerder de 'eigenlijke' manier van studeren, tegenover de 'oneigenlijke' manier. Op grond van de gekozen leeractiviteiten zijn de studenten in twee groepen gesplitst, de 'eigenlijke' en 'oneigenlijke' studeerders. De laatste groep blijkt in beide cursussen groter dan de eerste (bij Sociale Psychologie: 60%, bij Statistiek 1: 53%), hetgeen na het voorafgaande niet zal verbazen. In de tabellen 2a en 2b is de uitsplitsing gemaakt naar voorbereidingswijzen voor de gezakten en de geslaagden.

Tabel 2a: Aard van de voorbereiding gezakten en geslaagden Sociale Psychologie: aantal en percentage studenten dat gekozen heeft voor oneigenlijke resp. eigenlijke wijzen van voorbereiden.

	Oneigenlijk		Eigenlijk		Totaal
	N	%	N	%	
Gezakten	65	76	21	24	86
Geslaagden	49	48	54	52	103
Totaal	114		75		189

Tabel 2b: Aard van de voorbereiding gezakten en geslaagden Statistiek 1: aantal en percentage studenten dat gekozen heeft voor oneigenlijke resp. eigenlijke wijzen van voorbereiden.

	Oneigenlijk		Eigenlijk		Totaal
	N	%	N	%	
Gezakten	106	58	76	42	182
Geslaagden	30	39	47	61	77
Totaal	136		123		259

Van de studenten die gezakt zijn heeft 76% (Sociale Psychologie) respectievelijk 58% (Statistiek 1) zich voorbereid door vooral gebruik te maken van andermans uittreksels en van de geslaagde studenten is dat nog steeds het geval bij 48% (Sociale Psychologie) respectievelijk 39% (Statistiek1).

DISCUSSIE

Eerder is gesteld dat er goede grond is om het maximaal toelaatbare zakpercentage in het compromismodel te verhogen, als blijkt dat de zakkers (veel) minder dan de nominale studielast aan de voorbereiding van vakken besteden, of als zou blijken dat de aard van de voorbereiding bij een groot percentage zakkers te wensen overlaat. Een vergelijkbare constatering voor de geslaagden zou verhoging van de minimale kennisrechten rechtvaardigen.

De zakkers hebben met gemiddelden van 33 uur (Sociale Psychologie) en 50 uur (Statistiek 1) aanzienlijk minder gestudeerd dan de nominale studielast van 120 uur. Bovendien heeft 76% (Sociale Psychologie) en 58% (Statistiek 1) van de zakkers zich op oneigenlijke wijze op het tentamen voorbereid. De geslaagden halen met gemiddelden van 48 uur (Sociale Psychologie) en 62 uur (Statistiek 1) ook amper meer dan de helft van de nominale tijd. Nog altijd 48% (Sociale Psychologie) en 39% (Statistiek 1) van de geslaagden heeft zich op oneigenlijke wijze op het tentamen voorbereid. Bovendien blijkt dat de geslaagden bij Statistiek 1 na drie collegeweken gemiddeld slechts 35% van de stof hadden doorgenomen en aan het eind van de collegereeks, enkele dagen voor het tentamen, gemiddeld nog maar 71%.

Dat de discrepantie tussen norm en feitelijk gedrag groot is konden we weten uit het Sociaal en Cultureel Rapport 1998, dat gaat over de tijd die de gemiddelde Nederlandse student aan zijn studie besteedt, maar dat het gat zo gapend is heeft mij niettemin verbaasd. Alle redenen voor Examencommissies om zich in hun beleid niet meer naar de resultaten van dat feitelijk gedrag te richten? Alle redenen voor met name de Examencommissie Psychologie RuG om de kenniseisen fors aan te scherpen?

Relatieve methoden van cesuurbepaling – en in het voetspoor daarvan, de relatieve component van het compromismodel (de Z-max parameter) – zijn bedoeld om voor fluctuaties in de moeilijkheidsgraad van tentamens te corrigeren, waarbij ervan wordt uitgegaan dat studenten zich optimaal inspannen om aan de gestelde eisen te voldoen. Deze vooronderstelling wordt in ons huidige toetsstelsel blijkbaar niet waargemaakt, zeker niet in de onderbouw van grote faculteiten, waar tentamens in principe onbeperkt herkanst mogen worden en daarmee uitnodigen tot strategisch verkengedrag. Slechte toetsresultaten zullen dan ook eerder het gevolg zijn van onvoldoende studie-inspanning dan van onredelijk moeilijke tentamens. Daarmee vervalt overigens niet de zin om voor de invloed van (te) grote fluctuaties in moeilijkheidsgraad op tentamenresultaten te corrigeren. Het zakpercentage is daarvoor in ieder geval een ongeschikt middel gebleken.

Ook de constatering dat geslaagden voor propedeusetentamens naar eigen zeggen aanzienlijk minder tijd aan de voorbereiding besteden vormt de facto een pleidooi voor een meer absolute methode van cesuurbepaling, dus een cesuur die los staat van de toetsprestaties van studenten. Tevens is zonneklaar geworden dat de feitelijke time-on-

task voldoende ruimte laat voor een aanscherping van de kenniseisen.

Overigens blijkt uit dit onderzoek slechts dat de tentamenvoorbereiding van studenten voor verbetering vatbaar is, en *niet* dat de huidige voorbereidingswijze er de oorzaak van is dat de feitelijke cesuur amper boven het minimaal vereiste kennispercentage (score gecorrigeerd voor gissen) uitkomt. Het onderzoek is immers correlatief en laat zich derhalve niet causaal interpreteren. Het eerder aangehaalde onderzoek van Schouwenburg (1996), waaruit blijkt dat de hoeveelheid bestede tijd een belangrijke predictor is van studiesucces, maakt het echter plausibel dat een grotere tijdsinvestering leidt tot een betere stofbeheersing. Derhalve mag worden aangenomen dat een tijdig aangekondigde aanscherping van de cesuur (of een verhoging van de moeilijkheidsgraad van tentamens of een verlaging van de koers van studiepunten), door studenten opgevangen kan worden door meer tijd in de voorbereiding te steken. Althans door de meeste studenten, want voor studenten zonder een VWO-vooropleiding bleek in het geval van Statistiek 1 een relatief grote studielast samen te gaan met een zeer lage succeskans.

Tot dusver is de verklaring voor de discrepantie tussen norm en feit voornamelijk bij (het tijdbestedingspatroon van) studenten gelegd. Dat is natuurlijk maar het halve verhaal. Rationeel bestuurde onderwijsorganisaties zijn immers heel goed in staat om het rendement van hun inspanningen te verhogen door de time-on-task van hun studenten positief te beïnvloeden, bijvoorbeeld via opwaartse herzieningen van toelatings-, tentamenregelings- en cesuurparameters. De huidige beleidspraktijk stemt echter niet vrolijk. Veel facultaire bestuurders verwachten nog alle heil van onderwijsinnovaties waarvan we al lang weten dat ze vooral gunstig uitpakken voor studenten die zich ook zonder wel gered zouden hebben (teaching-learning paradox; Sipe & Curlette, 1996). Bovendien ontbreekt het diezelfde bestuurders aan moed of verbeeldingskracht om de toetsing als aangrijpingspunt te kiezen voor de kwaliteitsverbetering van leerresultaten. Meestal aan beide.

LITERATUUR

- Anderson, L.W. (Ed.) (1997). Educational testing and assessment: lessons from the past, directions for the future. *International Journal of Educational Research*, 27, 355-445.
- Boud, D. (Ed.) (1995). *Enhancing learning through self-assessment* (2nd edition). Londen: Kogan Page.
- Cohen-Schotanus, J. (1994). *Effecten van curriculumveranderingen. Studiewaardering, studeergedrag, doorstroom in een veranderend medisch curriculum*. Groningen: dissertatie RuG.
- Gibbs, G. (1999). Using assessment strategically to change the way students learn. In: Brown, S. & Glasner, A. (Ed.), *Assessment matters in higher education. Choosing and using diverse approaches*. Buckingham: Open University Press, 41-54.
- Hofstee, W.K.B. (1999a). Ritualisering van het onderwijs? *Tijdschrift voor Hoger Onderwijs*, 17, 20-30.
- Hofstee, W.K.B. (1999b). *Principes van beoordeling*. Lisse: Swets & Zeitlinger.
- Marton, F. & Saljö, R. (1976). On qualitative differences in learning II-Outcome as a function of the learners conception of the task. *British Journal of Educational Psychology*, 46, 115-127.

- McKenzie, G. (1973). Quizzes: Tools or tyrants. *Instruction Science*, 2, 281-294.
- Schouwenburg, H.C. (1996). Een onderzoek naar leerstijlen. *Tijdschrift voor Onderwijsresearch*, 21, 151-161.
- Sipe, Th. A. & Curlette, W. L. (1996). A meta-synthesis of factors related to educational achievement: a methodological approach to summarizing and synthesizing meta-analysis. *International Journal of Educational Research*, 25, 583-696.
- Smit, G. N. (1995). *De beoordeling van professionele gespreksvaardigheden. Constructie en evaluatie van rollenspel-, video- en schriftelijke toetsen*. Baarn: Uitgeverij Nelissen.
- Starren, J. (1998). Het mondeling tentamen als hefboom voor meer en beter leren. *Onderzoek van Onderwijs*, 27, 42-45.
- Vermunt, J. D. H.M. (1992). *Leerstijlen en sturen van leerprocessen in het hoger onderwijs: naar procesgerichte instructie in zelfstandig denken*. Lisse: Swets & Zeitlinger.
- Zelfstudierapport Psychologie RuG (2000). *Rapportage ten behoeve van de onderwijsvisietatie 2000*. Groningen: Intern rapport (Psych/ow/00.35).