

DE GRENS ZAKKEN / SLAGEN BIJ AFSLUITENDE TENTAMENS.

1.1 De vraagstelling is : "Hoe kan de grens tussen voldoende en onvoldoende in het onderwijs zo verantwoord mogelijk worden gesteld" (Wijnen, blz 9).

Deze vraagstelling is op twee manieren op te vatten, corresponderend aan verschillende onderwijs situaties. Het is mogelijk de waardering "voldoende" of "onvoldoende" te koppelen aan de beslissing "slagen" of "zakken", zoals in het onderwijs veelal gebeurt bij afsluitende toetsen. Een andere mogelijkheid is de vraagstelling op te vatten in deze zin: Hoe kan zo onduidelijk mogelijk aangegeven worden naar welk niveau van beheersing van de onderwezen stof gestreefd wordt? en: Hoe kan de beoordeling in de onderwijs situatie hier verantwoord aan gekoppeld worden ?. Laatst genoemde problematiek lijkt vooral van belang voor "voortgangstoetsen", die informatie moeten verschaffen over welke onderwijsactiviteiten het meest aangewezen zijn voor het bereiken van de omschreven onderwijs doelen.

1.2 De methode Wijnen voor de cesuur bepaling is alleen relevant voor de afsluitende toetsing; uit de aard van de methode volgt dat ze geen bijdrage kan leveren wanneer studietoetsen als "diagnostisch" hulpmiddel in het onderwijs gehanteerd worden (zoals boven kort omschreven). hiermee is tevens een studiethema geïntroduceerd waaraan in het volgende voortdurend aandacht zal worden besteed:

Welke is de relatie tussen de afsluitende toets en het eraan voorafgaande (en het erop volgende) onderwijs, en welke zijn daarvan de consequenties ten aanzien van de interpretaties die aan de toetsingsresultaten gegeven kunnen of mogen worden?

1.3 De selectieve functie van de toets.

Met selectieve functie wordt bedoeld dat de toets informatie over studenten levert, op basis waarvan uitspraken gedaan worden over toelating tot volgende studie fasen. De veronderstelling daarbij is veelal dat het betreffende studie onderdeel met goed gevolg afgesloten moet zijn, wil de student het verdere onderwijs met vrucht kunnen volgen (een "noodzakelijke voorwaarde") ; óf er kan een relatie verondersteld worden in de zin dat deze

toets resultaten verband houden met later "studiesucces" (er wordt dan gesproken van "voorspellende geldigheid" van de toets voor uiteindelijk studie succes).

Wanneer men aan een studietoets een dergelijke selectieve functie wil toekennen, is een behoorlijke bewijsvoering met betrekking tot de juistheid van het veronderstelde karakter van "noodzakelijke voorwaarde" en/of van de voorspellende geldigheid, gewenst. Beide problemen zullen in het volgende niet zozeer in de aandacht staan. omdat ons primair interesseert hoe de relatie van de beoordeling, en in het bijzonder de zak-slaag beslissing, is met het eraan voorafgaande onderwijs en de verwachtingen die er over de resultaten van dat onderwijs bestaan bij de docenten.

1.4 het doel van het toetsgebruik.

Vrijwel zonder uitzondering zijn studietoetsen bedoeld om op enigerlei wijze te meten in welke mate de individuele student of een groep van studenten net onderwezene "beheersen". Met deze korte formulering gaan we voorlopig geheel voorbij aan de problemen die zich praktisch en theoretisch bij dergelijke metingen voordoen.

Een dergelijke vorderingenmeting impliceert op geen enkele wijze de noodzaak om gebruik te maken van cijfers, of om predikaten als "voldoende", "onvoldoende" etc te hanteren. Er moet een duidelijk onderscheid gemaakt worden tussen de meting op zich, en het gebruik dat men ervan wil maken, zij het ook met de belangrijke restrictie dat de meting afgestemd moet worden op dat gebruik dat men ervan gaat maken. Dit laatste punt behoeft enige toelichting:

Wanneer men een toets afneemt om vorderingen te meten op een bepaald stuk onderwezen stof, of men neemt een toets af om op grond van de resultaten beslissingen te nemen over het al dan niet toelaten tot verder onderwijs, dan moeten in beide gevallen heel andere eisen aan de te construeren toets gesteld worden. De betekenis hiervan voor onze probleemstelling is, dat resultaten op een vorderingen toets niet zonder meer gebruikt kunnen worden om er bepaalde beslissingen, zoals de zak-slaag beslissing, aan te verbinden.

1.5 Het begrip cesuur. Schets van een methode.

Omdat we het cesuurprobleem bestuderen zoals zich dat voordoet bij de eindtoets voor een enkel studieonderdeel, is er slechts een cesuurpunt, nl. de

grens tussen voldoende en onvoldoende. Wanneer meer mogelijkheden dan deze twee toegelaten worden hebben we veeleer te maken met een voortgangstoets (zie 1.1).

Een mogelijke strategie voor het vinden van de "beste" cesuur is, om allereerst na te gaan waar de scheiding tussen onvoldoende en voldoende zou moeten komen te liggen wanneer we alleen letten op de inhoud van het onderwezene, en de gekozen doelstellingen m.b.t. het gewenste niveau van beheersing over de stof. De tweede stap zou zijn de zorgvuldige afweging van de consequenties die de te nemen beslissingen "voldoende" of niet, hebben voor zowel de individuele student als voor de onderwijsinstelling (vanwege de gevolgen voor de kwaliteit en de organisatie van het onderwijs die verschillende selectiestrategieën hebben). Overwegingen gebaseerd op deze tweede stap zullen verschuivingen teweeg brengen in de aanvankelijk gekozen cesuur. Beide genoemde fasen zijn echter enigszins abstract, zodat nog een derde fase nodig is waarin de genoemde overwegingen gerelateerd worden aan de te gebruiken eindtoets. Daarbij worden zowel de kwaliteit van de gebruikte toets, als de resultaten die er op behaald worden, gebruikt om aan de eerdere, meer abstracte overwegingen, een concretisering te geven die leidt tot een weloverwogen cesuur bepaling.

Wanneer de geschetste methode gevolgd wordt, behoeven we ons geen zorgen te maken over de vraag of de grensscore van tevoren of pas achteraf (als de feitelijke resultaten bekend zijn) vastgesteld moet worden. Het gekozen schema impliceert immers een vooraf gekozen strategie, waarbij gebruik gemaakt wordt van de feitelijk bereikte resultaten om een bevredigende correctie op een van tevoren vastgestelde grensscore aan te brengen. De controverse die bestaat tussen pleiters voor het systeem van absoluut meten en degenen die voorstander zijn van "relatief" meten in het onderwijs, lost zich hopelijk op in een methode die de voordelen van beide systemen in zich kan verenigen zonder met de nadelen behept te raken.

2. De eerste stap: bepaling van verlangd niveau van beheersing over de stof.

2.1 De doelstellingentabel.

Een uitstekend hulpmiddel bij deze eerste stap is een doelstellingen tabel. In een dergelijke tabel zetten we horizontaal de verschillende onderdelen van het onderwezen vak af, en verticaal de indeling van een doelstellingencategorieënsysteem, waarvoor we bijvoorbeeld de hoofdcategorieën van het systeem van Bloom kunnen nemen. Door deze

horizontale en vertical indeling zijn een groot aantal cellen in de tabel bepaald, waarvan we na moeten gaan of er voor ons relevante doelstellingen in passen. Voorbeelden van dergelijke doelstellingen tabellen zijn te vinden in W. Meuwese, onderwijs research, blz. 56; in Bloom, Hastings and Madaus: Handbook on formative and summative evaluation of student learning, worden voorbeelden van constructie van dergelijke doelstellingentabellen gegeven, naast voorbeelden uit de concrete onderwijspraktijk waarin met dergelijke tabellen gewerkt is.. Een dergelijke tabel kan ons op velerlei manieren van pas komen, bijvoorbeeld als controle op de volledigheid van een voor een bepaald tentamen gebruikte set vragen. Wanneer we deze vragen thuis brengen in de doelstellingen tabel zal vaak blijken dat bepaalde cellen leeg blijven, terwijl uit andere cellen veel vragen gesteld worden, zonder dat een en ander in overeenstemming is met de meest door ons gewenste verdeling over de cellen.

2.2 Omschrijving van nog juist voldoende niveau van beheersing.

Is een doelstellingentabel eenmaal geformuleerd, dan kan aan de hand van deze tabel een poging ondernomen worden om een precieze omschrijving te geven van de beheersing over de stof die iemand moet tonen om nog juist als "voldoende" beoordeeld te kunnen worden.

Wanneer net de bedoeling is om de stof te toetsen met behulp van meerkeuze items, zou de volgende werkwijze (bijvoorbeeld) gevolgd kunnen worden. Allereerst wordt gezocht naar een representatie van de cellen van de doelstellingen tabel in aantallen items die betrekking moeten hebben op iedere afzonderlijke cel. Op deze wijze wordt een voor iedereen (ook studenten!) inzichtelijk overzicht van de totale leerstof verkregen, en van het belang dat diverse onderdelen van die stof in de eindtoets krijgen. Lukt een en ander op bevredigende wijze, dan is er al een niet geringe prestatie geleverd op de weg naar het doorzichtig maken van de toetsingsprocedure.

Aan de hand van de doelstellingentabel met "celfrequenties" of "celgewichten" kunnen we vervolgens proberen een omschrijving te geven van het niveau van de "minimum voldoende" prestatie. Eén van de mogelijkheden daarbij is om te volstaan met het aangeven van een percentage van de vragen dat goed gemaakt moet worden (daarbij kan voorlopig ook nog afgezien worden van de invloed van raadmogelijkheden e.d., deze kunnen immers in een later stadium nog ingecalculeerd worden)

Het is misschien wel handig om daarbij uit te gaan van een denkbeeldige toets bestaande uit (bijv.) 1000 items, zodat het ook bij grote aantallen cellen in de doelstellingentabel mogelijk is om verschillen in belangrijkheid van de

stof in verschillende cellen tot uitdrukking te brengen in een verschillend aantal vragen over die stof te stellen. Bij de samenstelling van een concrete toets van bijv. 60 items moet er dan voor gezorgd worden dat deze 60 een reële afspiegeling van de oorspronkelijke 1000 items vormen.

Naast deze aller eenvoudigste procedure zijn er nog vele andere mogelijke werkwijzen, zoals het hanteren van de methode Nedelsky of de kernitem-methode voorzover zij in dit stadium relevant kunnen zijn. Zie Meuwese, blz 180, en het boek "studietoetsen, construeren, afnemen, analyseren" van De Groot, van Naerssen e.a. op de blz 278 en 273. Misschien is het ook mogelijk om op enigerlei wijze te komen tot een minder zware weging voor kennisvragen dan voor bijv. inzichtvragen. Deze werkwijze stuit op een complicatie, die overigens ook aan de andere methoden inherent is:

Eenzelfde eindscore kan op vele uiteenlopende manieren bereikt worden (afgezien van de rol die raadmogelijkheden spelen). Dit is een gebrek dat aan iedere methode van cijfer toekenning kleeft, en dat bewust geaccepteerd zal moeten worden zolang men op enigerlei wijze aan het traditionele cijfersysteem wil vasthouden. Dat neemt niet weg dat men wel globaal kan analyseren op welke verschillende manieren een als juist voldoende aangemerkte eindscore bereikt kan worden, en dat bij nader inzien als niet aanvaardbaar beschouwde scorepatronen waarschijnlijk uitgeschakeld kunnen worden door in de doelstellingentabel de vragenfrequenties in de cellen te veranderen. De gebleken discrepantie tussen de eerste doelstellingenformulering en de eisen die in de beoordelingssituatie gesteld worden proberen we door de omschreven terug koppeling van gegevens teniet te doen. De discrepantie wordt dus opgeheven ten gunste van de beoordelingssituatie. Wanneer dit gebeurd is geeft de doelstellingen tabel niet meer het belang van ieder onderdeel van de stof t.o.v. andere onderdelen aan, maar de tabel is nu geheel toegespitst op de beoordelingssituatie, in die zin dat zij aangeeft welke "wegen" naar een nog juist voldoende eindscore kunnen en mogen leiden.

Het voorgaande is uiteraard niet meer dan een schets. In een poging tot concretisering van een en ander zal men op vele problemen stuiten, die meestal door onderlinge discussie en door expliciete analyse van de doelen die men met een bepaald stuk onderwijs nastreeft, opgelost kunnen worden. Eén van de zaken die in het bovenstaande onaangeroerd zijn gebleven is de kwestie van de moeilijkheid van de te hanteren toetsitems. Natuurlijk moet met een gegeven als de verwachte of de gebleken moeilijkheid van de gestelde vragen rekening gehouden worden bij iedere cesuurbepaling op basis van dergelijke vragen gedaan. Dit kan op verschillende manieren: het probleem uitstellen totdat de toets afgenomen is (dus tot de derde stap in de cesuurbepalingsprocedure); ofwel eerst het gegeven model uitwerken, daar

meerkeuzevragen bij formuleren, de moeilijkheid van iedere vraag schatten, en vervolgens nagaan of er veranderingen in de doelstellingen tabel aangebracht moeten worden op grond van de extra informatie over de verwachte moeilijkheid van de vragen;. Dit soort problemen zijn niet dermate complex dat in dit paper op al dergelijk soort moeilijkheden geanticipeerd moet worden met mogelijke oplossingsmethoden of werkwijzen. (nog iets over de moeilijkheid van studietoets items: Het blijkt dat schattingen door docenten van de moeilijkheid van de items niet goed mogelijk zijn. Dit is dan de moeilijkheid in termen van het percentage studenten dat de vraag goed blijkt te kunnen beantwoorden. Natuurlijk kan moeilijkheid ook opgevat worden in de zin : de kans dat iemand die een nog juist voldoende niveau van beheersing over de stof heeft een bepaald item goed kan maken. Deze definitie, of een andere uit het grote scala van mogelijkheden, kan dan als voorlopige leidraad genomen worden, totdat in de derde stap concrete gegevens over de gebleken moeilijkheid van de items beschikbaar komen. Op basis van de feitelijke informatie kan dan bezien worden of in de oorspronkelijke moeilijkheidsbeoordelingen wijzigingen aangebracht moeten worden. Natuurlijk kan dit soort informatie ook gebruikt worden om te evalueren hoe het door de docent verzorgde onderwijs bij de studenten is overgekomen, op welke punten de verwachtingen van de docent overtrokken waren, etc. Wanneer met het volgen van de geschetste procedure enige ervaring is opgedaan kunnen nu nog veronderstellenderwijs ondernomen activiteiten op grond van de dan verkregen feedback afgekapt of voortgezet worden, in al dan niet verbeterde versie.)

2.3 De gehanteerde onderwijsfilosofie.

Het is niet de bedoeling hier een beschouwing over de verschillende opvattingen over aard en inrichting van het onderwijs te geven. Slechts wordt er op gewezen dat aan het onderwijs, waaraan een cesuurproblematiek, zoals hier omschreven in belangrijke mate inherent is, een bepaalde onderwijsopvatting ten grondslag ligt. Het is in principe mogelijk om het onderwijs dusdanig vorm te geven dat het cesuurprobleem daarin een ondergeschikte plaats kan innemen, in tegenstelling tot de traditionele praktijk. In plaats van de als godsoordeel gebruikte eindtoets kan er al tijdens het onderwijs frequent getoetst worden, met direct er op volgende herkansingsmogelijkheden, waarbij de student gebruik kan maken van een goede analyse van de tekortkomingen op de in eerste termijn door hem niet bevredigend gemaakte deelttoets. Wanneer een dergelijke onderwijsvorm gerealiseerd wordt moet het mogelijk zijn dat iedereen (vrijwel iedereen) binnen redelijke termijn een bepaald stuk onderwijs met voldoende resultaat kan afsluiten. Een afsluitende toets, die dan bijv. om onderwijskundige redenen nog gehouden zou kunnen worden, zou dan ook door de goede

voorbereiding in de onderwijsfase door vrijwel iedereen voldoende gemaakt kunnen worden (voldoende dan in de zin van voldoende aan de minimum eisen die aan beheersing van de onderwezen stof gesteld worden).

Met dergelijke vormen van onderwijs wordt o.a. door Bloom geëxperimenteerd, zie ook de talrijke verslagen in het reeds geciteerde boek van hem. Ook in Nederland wordt met dergelijke onderwijs vormen al geëxperimenteerd, al dan niet expliciet onder het label van het door Bloom e.a. gepropageerde "learning for mastery" model.

Hoewel het van het grootste belang is om voor de traditionele afsluitende toets tot een verantwoorde cesuurbepaling te komen, moet daar toch de vraag bij gesteld worden of al deze moeite niet beter gericht zou kunnen worden op het realiseren van vormen van onderwijs waarin de beoordelingsmomenten volkomen in het onderwijs en leerproces zelf geïntegreerd zijn.

3. De "gevolgen" van de gekozen zakslaag grens voor onderwijs en student

Gesteld wordt de vraag naar de bruikbaarheid van studietoets en toetsresultaat voor de beslissingen die op basis van deze informatie genomen moeten worden.

De toetsresultaten vormen de informatie op grond waarvan de faculteit (i.c. de docent) zakslaag beslissingen neemt. Daarbij moeten we niet uit het oog verliezen dat ook de student op basis van zijn toetsresultaten zal beslissen om al dan niet door te gaan met de studie, bepaalde specialisaties te kiezen, etc. Ook voor deze door de student te nemen beslissingen draagt de faculteit verantwoordelijkheid, in deze zin, dat zij er voor moet zorgen dat de informatie die de toets oplevert ook relevant is voor deze door de student te nemen beslissingen omtrent zijn studie. Het is inderdaad de vraag of één en dezelfde toets wel informatie op kan leveren die voor beide vraagstellingen de meest relevante is. In de mate waarin dit niet het geval is, zal daarin door andere informatie verzamelingsmethoden voorzien moeten worden.

Cronbach en Gleser (in het boek Psychological Tests and Personnel Decisions) behandelen uitgebreid het probleem van selectie op basis van univariate (of univariaat gemaakte) informatie, waar alle geslaagde personen een gelijke "behandeling" in het verdere onderwijs te wachten staat. Voor het cesuur probleem is echter de behandeling die de geslaagde studenten in het verdere onderwijs krijgen niet van direct belang, wèl de behandeling die de gezakten krijgen. Vanuit de docent gezien is de

behandeling van de gezakten gelijk: ze moeten allen een stuk onderwijs of zelfstudie overdoen. Vanuit de individuele student bezien zijn de behandelingen echter zeer ongelijk, in termen van de door hem in te lopen achterstand, zijn studiemotivatie, zijn persoonlijke en materiële omstandigheden voorzover relevant voor en/of beïnvloed door de consequenties die het gezakt zijn voor hem heeft.

Evenals Cronbach en Gleser zullen we voornamelijk aandacht schenken aan de zgn. "institutionele" effecten van de beslissing tot zakken, en de gesignaleerde consequenties voor de individuele student (tenminste voorlopig) niet tot kern van het betoog maken. Ook hier moet weer gesignaleerd worden dat deze benadering van de problematiek het traditionele, selectieve, beoordelingsstelsel veronderstelt; bij meer onderwijskundig en anderszins verantwoorde onderwijsprogrammering verandert de selectieproblematiek in een sturingsproblematiek, waarvoor vele van de in dit hoofdstuk vermelde overwegingen minder of in het geheel niet relevant zijn.

We nemen dus aan dat voor alle gezakten geldt dat ze een stuk onderwijs (of zelf studie) moeten overdoen. Bovendien nemen we aan dat op geen enkele wijze aantallen of percentages gezakten van te voren vastgelegd zijn (wanneer dit wel het geval is, bij een al dan niet verkapte numerus fixus bijvoorbeeld, heeft dit weer heel specifieke consequenties voor de dan te volgen beslissingsprocedure).

Een beslissing tot zakken is te zien als een besluit om deze student aan een nieuwe toets te onderwerpen, om op basis van deze meerdere informatie (of hij in staat geweest is om zich in tussentijd alsnog tot het verlangde niveau op te werken) opnieuw de zak slaag beslissing in overweging te nemen. Iedere beslissing tot zakken is, gezien vanuit de onderwijs instelling, in feite een opschorting van de eindbeslissing: slagen of consilium abeundi. Ook de student kan het zo zien: haal ik het niet de eerste keer, dan toch zeker de tweede keer. In welk geval de student het spelletje meespeelt, door heel bewust een bepaald zakrisico te nemen, in ruil voor een besparing in tijd en energie. Het is dan wel de vraag (1) of dit een in het onderwijs wel zo gewenste instelling is, en (2) of een dergelijke strategie inderdaad door een belangrijk deel van de studenten gekozen is. Wanneer bijvoorbeeld een tentamen werkelijk te zware eiden stelt, dan blijft er voor heel weinig studenten de speelruimte over die nodig is voor het kunnen volgen van deze strategie van het bewust geaccepteerde risico. In dat geval zullen de consequenties van de beslissing tot zakken zowel voor de onderwijsinstelling als voor de individuele student, veel ernstiger zijn. De gezakte student is nauwelijks in staat zich op een herkansing beter voor te bereiden dan op de eerste keer. Voor de onderwijsinstelling betekent het dat de studieduur

onnodig verlengd wordt, en daarmee tevens de kosten van het onderwijs (en de maatschappelijke kosten). Zeer groot gewicht komt in dat geval toe aan de argumentatie die gevoerd wordt voor het hanteren van een zo hoge selectiedrempel.

Voor de gezakte student is de uiteindelijke beslissing dus uitgesteld. Van groot belang is derhalve, wat er in de tussentijd gebeurt met betrekking tot de onderwijsdiensten die deze student aangeboden worden. In het traditionele tentamensysteem wordt de gezakte student nauwelijks begeleid om tot een alsnog voldoende prestatie te komen. Ook dat betekent weer dat er een vrij grote onzekerheid moet bestaan (ook al is de docent misschien niet zo geneigd dit onmiddellijk in te zien) over de waarde van de voorlopig afwijzende beslissing, omdat de onderwijskundig te waarderen gevolgen van die beslissing moeilijk te overzien zijn,. Met andere woorden, hoe beter de voorlopig afgewezen student door de docent of de faculteit opgevangen en begeleid wordt, des te duidelijker zijn de consequenties van de zakslaag beslissing, en met des te groter vertrouwen kan de docent dergelijke beslissingen nemen, terwijl ze in die mate ook acceptabeler zullen zijn voor de er bij betrokken studenten.