

<p><br>KOMPENSATORISCHE EXAMENREGELING

<p><br>De kompensatorische examenregeling vraagt de student om <i>gemiddeld</i> over alle studieonderdelen aan een minimum eis te voldoen. Dit in contrast met de conjunctieve examenregeling waarbij de student voor ieder studieonderdeel afzonderlijk aan een minimum eis voldaan moet hebben . Konjunctieve regelingen zouden onwerkbaar zijn wanneer geen herkansingsmogelijkheden geboden zouden worden. De kompensatorische regeling kan in beginsel z&acute;nder dergelijke herkansingen functioneren.

<p>

Het moet dan ook mogelijk zijn om aannemelijk te maken dat onder kompensatorische regelingen dezelfde resultaten te bereiken zijn als onder conjunctieve regelingen, maar in minder tijd. Een belangrijk deel van de tijd die door studenten aan het voorbereiden van, en afleggen van, herkansingen wordt besteed onder conjunctieve regelingen is namelijk onproductief bestede tijd.<p><br>

Een monografie over deze problematiek wordt eind 1980, begin 1981 geschreven, en zal de hoofdmoot vormen van een academisch proefschrift. Als promotor is bereid op te treden prof. dr. W.K.B. Hofstee, maar mede op zijn suggestie zal nog nagegaan worden of het niet beter is of meer voor de hand ligt om prof. dr. R.F. van Naerssen te verzoeken als promotor op te willen treden. <p><br>

De idee voor deze verhandeling is al oud. Mijn hoogleraar Dijkhuis (Utrecht) was gewoon met betrekking tot herkansingen voor tentamens de uitspraak te doen dat nimmer was aangetoond dat dergelijke herkansingen zinvol waren, de student een betere greep op de stof gaven dan hij/zij bij eerste opkomst al had. <p><br>

Een eerste konkretisering van deze problematiek heb ik gegeven in een invited paper 13-12-74 voor een studiedag van de economische fakulteit van de R.U. Groningen. Dit paper was een uitwerking van de uitspraak van Cronbach en Gleser (1957) dat aan verschillende beslissingen (zakken of slagen) over personen de veronderstelling van het bestaan van een Aptitude Treatment Interaction ten grond ligt. <p>

In een invited paper voor een RWO 'methodologie professionaliserings dag' (6 november 1975) werd e.e.a. een stapje verder uitgewerkt. In augustus 1976 werd 'Argumenten voor afschaffen van het overdoen' geschreven, in dec. '76 toegespitst tot een 'Voorlopige inventarisatie van onderwerpen voor een Kleine logica voor het niet overdoen van onvoldoende gedane vakken.' Het waren deze stukken die Hofstee er toe brachten mij aan te sporen dit tot een proefschrift uit te werken. <p><br>

November 1977 werd een onderzoekvoorstel bij de CVBHO (Cie Buijs) ingediend, over Effect van

verschillende examenregelingen op doorstroming en studierendement. Van de Cie. is nimmer enige reaktie vernomen. <p><br>

Laatstelijk werd in een invited paper voor de ORD 1979 een kort overzicht van deze examenproblematiek gegeven: 'Universitaire examenregeling: conjunctief of kompensatorisch? <p><br>

Het werk aan cesuurbepaling en tentamenmodellen houdt nauw verband met de examenproblematiek, en levert vele inzichten en technieken die onmisbaar zijn bij het vnl. analytisch demonstreren van bet waarschijnlijk spektakulaire overwicht van kompensatorische op conjunctieve regelingen waar het om het studierendement, en dus om de kosten van het onderwijs gaat. Daar valt nog aan toe te voegen dat zeker niet alleen in het wetenschappelijk onderwijs baten zijn te verwachten van een accentverschuiving van conjunctieve naar kompensatorische examenregelingen. <p><br>

In hoeverre gaat het hier nu om ideeën die oorspronkelijk zijn? Tenslotte hebben discussies over de verdiensten van conjunctieve, kompensatorische, en zelfs disjunctieve examenregelingen wel eerder plaats gevonden, en bestaan er met name ook wel propedeuse-regelingen die sterke kompensatorische trekken hebben. Het gaat dan ook niet zozeer om de propositie zelf, als wel om de aard en de kracht van de argumenten die tot haar

ondersteuning worden aangevoerd. Het gaat om het bijeenbrengen van verschillende argumenten, zoals de al genoemde ATI veronderstelling, en het vervolgens in een zorgvuldige (modelmatige) analyse demonstrenen dat de ene soort examenregeling vraagt om een stuk onproductieve tijdbesteding van de student die onder de andere soort regeling niet voorkomt, en dat terwijl andere essentiële onderwijsparameters gelijk gehouden worden. <p><br>

Dat het kompensatorisch inderdaad beter zou kunnen, is overigens een stelling die op voorhand door velen bestreden wordt, althans ontkend zal worden. Er is dan ook zeker geen sprake van een poging om het voor-de-hand-liggende te bewijzen. Mij is alleen van Van Naerssen een uitspraak bekend (1974, advies aan onderwijs cie. subfak. psychologie UvA) dat van alle beleidsalternatieven m.b.t. herkansingen het afschaffen ervan de meest effectieve is. <p><br>

## CESUURBEPALING<P><BR>

Een goede methode voor cesuurbepaling is &acute;&acute;n van de meest gezochte stenen der wijzen in de onderwijsresearch. Dat is al zo voor de traditionele toetsvormen, en dat is wel heel sterk het geval voor criterium-gerefereerde (of domein-gerefereerde) toetsvormen. Het streven is naar een methode die een cesuur oplevert die in omschreven opzicht optimaal is. <p><br>

De problemen bij het vaststellen van zak-slaaggrenzen zijn groot, en lijken vrijwel onoplosbaar te zijn. Ik heb dat uitgebreid beschreven in Cesuurbepaling (1e versie, mei 1977). Het opmerkelijke van deze als docentenkursus uitgegeven monografie was dat nu eens niet een bepaald 'ei van Columbus' gepresenteerd werd, een slim techniekje waardoor je cesuurprobleem 'opgelost' wordt, maar dat geprobeerd werd om vanuit de algemene onderwijsdoelstellingen redenerend een methode voor cesuurbepaling te konstrueren, een methode die dan ook congruent met die onderwijsdoelstellingen zou zijn. Deze konstruktiepoging slaagde in deze eerste versie niet, maar maakte wel duidelijk waar de knelpunten zaten. En dat waren knelpunten waar ook andere methoden niet omheen konden. <p><br>

Cesuurbepaling op basis van tests of toetsen is een klassifikatie-probleem. Enkele technische problemen die zich bij het klassificeren van personen voor kunnen doen werden al in 1968 onderzocht tijdens een stage bij de groep onderwijsresearch aan de T.H. Eindhoven: 'Multiple discriminant analyse van de Cattell 16 P.F. voor studenten in zeven studierichtingen aan de T.H. Eindhoven. Daar bleek al dat het onder 'normale' omstandigheden wel eens niet mogelijk kan zijn op goede wijze bepaalde klassen te vullen (die blijven dan leeg). Een verschijnsel dat zich ook bij zak-slaagbeslissingen voor kan doen (zie 'Optimale grensscores zijn eenvoudig te vinden' TOR 1980), of bij toelating tot

numerus fixus studies (Selectie voor numerus fixus is irrationeel', september 1974). <p><br>

De besliskundige benadering van de toelatingsproblematiek bij numerus fixus (in 'Overzicht van de selectieproblematiek bij numerus fixus regelingen,' april 1975) leidde als vanzelf tot eenzelfde aanpak van de cesuurproblematiek (zie de bijlage bij genoemde studie: 'Institutioneel selectiemodel voor selectie op geschiktheid'). Daarop volgden een aantal memo's over het nemen van beslissingen, uitmondend in het kursuskoncept van augustus 1976 waaruit de eerste gepubliceerde versie van mei 1977 ontwikkeld werd, Hoofdstuk 7 van 'Cesurbepaling' geeft een uiteenzetting van de besliskundige benadering van de cesuurproblematiek. <p><br>

Op zich is een besliskundige benadering geen 'vondst'. Van Naerssen had de noodzaak daarvan al vaker aangeduid, in de psychometrie waren al aanzetten te vinden, en met name voor criterium-gerefereerde toetsen kwamen juist de 'besliskundig georiënteerde' studies los. Alle auteurs echter baseerden zich op een besliskundige benadering rechtstreeks ontleend aan de statistische tak van de besliskunde: een korrekte maar omslachtige method bekend onder de naam 'normal form analysis'. Een even korrekte maar veel eenvoudiger methode is de 'extensive form analysis', en het is deze methode die ik van meet af aan gebruikte. (Zie ook de par. 'extensive en normal form analysis' in 'Enkele

radikale oplossingen voor criterium gerefereerde grensskores'). Vele besprekingen in de VOR - werkgroep criterium-gerefereerde toetsing hebben mij geleerd dat het allerm minst eenvoudig is voor degenen die de mathematisch gekompliceerde normal form analysis gehanteerd hebben om de eenvoudige extensive form analysis als een korrekte methode te herkennen. Hoe dat ook zij, hanteren van de extensive form analysis maakt een besliskundige aanpak van cesuuproblemen zeer veel eenvoudiger dan tot nu toe gegeven oplossingen. <p><br>

In het artikel 'Optimale criterium gerefereerde grensskores zijn eenvoudig te vinden' wordt de extensive form analysis gepresenteerd. Bij deze cesuurbepaling speelt de selektieparameter  $P$  een rol, de proportie 'geschikten' in de deelgroep met een test- of toetsskore gelijk aan de (optimale) cesuur. Aangetoond wordt dat in een aantal verschillende methoden uitgaande van de normal form analysis dezelfde parameter de (optimale) cesuur bepaalt, m. n. ook bij Huynh's (1976) aanpak. Hiermee wordt het helder dat de problematiek niet is om bij ieder aangedragen model weer opnieuw te bewijzen dat  $P$  de optimale cesuur bepaalt (tot nu toe de grootste zorg van auteurs), maar om modellen te ontwikkelen die tot goede (betere) schattingen voor de waarde van deze selektieparameter leiden. In de diskussie wordt er op gewezen dat de (impliciete) veronderstelling bij methoden van cesuurbepaling aangedragen vanuit

psychometrische hoek is dat er duidelijk kwalitatieve grenzen in de mate van stofbeheersing aanwijsbaar zouden zijn, een veronderstelling die niet houdbaar lijkt te zijn en ook niet nodig is voor het vinden van methoden voor cesuurbepaling. Deze simpele konstatering ligt aan de basis van een alternatieve, en hopelijk vruchtbare, benadering van het cesuurprobleem, zoals gedeeltelijk al uitgewerkt in het vervolgartikel 'Enkele radikale oplossingen etc.', en uit te werken in de tweede versie van Cesuurbepaling. <p><br>

In 'Enkele radikale oplossingen voor criterium gerefereerde grensscores' wordt een ernstige methodologische tekortkoming in alle tot nu toe gepresenteerde besliskundige methoden van cesuurbepaling blootgelegd, en gerepareerd. Impliciet werd het effect van 'bijspijkeren' of herkansen opgenomen in de toekenning van utiliteiten aan de vier mogelijke uitkomsten van doorlaten of afwijzen. Dit mankement is te repareren, hoewel de methoden er daardoor niet direkt eenvoudiger op zullen worden. In de nu gepresenteerde aanpak zijn utiliteiten niet meer gekontamineerd door (onbekende) waarschijnlijkheden, en wordt het tot nu toe onopgeloste probleem hoe je op een goede manier je utiliteiten moet bepalen, opgelost. De belangrijke dwarsverbinding met de Aptitude Treatment Interaction methodologie (Cronbach & Snow 1977) blijkt nu eenvoudig te leggen. Een aantal terminologische problemen die nogal eens

aanleiding gaven tot verwarring en onjuiste konklusie in de literatuur, zijn in dit artikel eveneens meegenomen. <p><br>

Nu de criterium gerefereerde cesuurbepaling hiermee, althans wat betreft de besliskundige benadering ervan, zijn juiste proporties gekregen lijkt te hebben, is ook beter in te zien dat in veel onderwijssituaties (bij Individuele Studie Systemen bijvoorbeeld) geheel andere methoden minstens even goed of misschien veel beter bruikbaar zijn, zoals in de discussie van het artikel aangegeven. En dat laatste geldt met name ook voor cesuurbepaling bij tentamens. <p><br>

Besliskundige benaderingen van het cesuurprobleem zijn vooral d&acute;n ge&iuml;ndiceerd wanneer er met aanzienlijke onzekerheden gewerkt moet worden. In lopend onderwijs zijn de onzekerheden tav. de groep studenten (of deelgroepen studenten) echter niet zo groot, en kun je beter gebruik maken van methoden waarbij op zijn hoogst een trade-off tussen enkele verschillende doelvariabelen zoals tijdbesteding, en bereikte stofbeheersing bepaald moet worden. Wat niet wegneemt dat je overigens bij het verzamelen en interpreteren van je data geen goede (statistische) procedures zou hoeven te volgen. <p><br>

Op het voorgaande wordt de tweede versie van Cesuurbepaling gebaseerd. Daarbij zal met name ook gebruik gemaakt worden van de belangrijkste

resultaten die het onderzoek naar tentamenmodellen (Examenregeling deel A: Studiestrategie&uml;n) heeft opgeleverd. Voor het eerst zal dan een methode voor cesuurbepaling beschikbaar zijn die in beginsel doorzichtig is, die evalueerbaar is, waarvan de werking bekend en toetsbaar is. Ook in deze nieuwe versie zal aan juridische aspecten van deze vorm van toetsing aandacht besteed worden, evenals aan de Richtlijnen en hun betekenis voor het nemen van zak-slaagbeslissingen. Deze beschouwing van de administratief rechtelijke aspecten van beoordelen in het onderwijs (algemene beginselen van behoorlijk bestuur), voor het eerst gegeven in de 1977 versie van Cesuurbepaling, is de eerste aanzet in Nederland voor het leggen van een dwars-verbinding tussen 'rechten' en 'didakometrie'.

## <p><br>TOETSEN<p><br>

Wat toetskonstruktie en -analyse betreft heb ik mij vooral bezig gehouden met de problematiek rond meerkeuzevragen, mogelijkheden voor het genereren van toetsvragen, betrouwbaarheid, en de interpretatie van toetsresultaten. <p>

In typisch Nederlands spraakgebruik is een 'studietoets' een toets die bestaat uit twee- of meerkeuzevragen. Vanaf de introductie in ons land van 'de meerkeuzevraag' heeft over de wenselijkheid daarvan een discussie bestaan, een discussie die

eind zestiger jaren in het voordeel van de 'voorstanders' beslecht leek, zonder dat overigens weerstanden in de samenleving tegen deze vraagsoort (met de 'CITOtoets' als exponent) geslecht konden worden. Mijn bezwaren tegen het klakkeloos en kritiekloos gebruiken van deze vraagsoort waren velerlei: het ontbreken van regels voor het inhoudelijk verantwoord konstrueren van dergelijke vragen (ik kom daar straks uitgebreider op terug); misbruik van het objectiviteits-argument; vaak minder gelukkig gekozen aantal alternatieven; slechte onderwijskundige kwaliteit van nogal wat van dergelijke vragen; het nogal overtrokken voordeel van de automatische skoorbaarheid. <p><br>

In de Amerikaanse literatuur is het gebruikelijk zowel kort-antwoordvragen en invulvragen als meerkeuzevragen tot de objektieve vragen te rekenen. In ons land heeft De Groot benadrukt dat een uniek voordeel van de meerkeuzevraag haar objektieve skoorbaarheid is, waardoor het ook de enige vraagsoort zou zijn die een faire beoordeling van de leerling (student) mogelijk maakt. Ik heb dit Grootiaanse standpunt bestreden in een ORD paper (Het verborgen vooroordeel tegen andere dan meerkeuze vraagvormen, 1977), erop wijzend dat objektieve skoorbaarheid niet verheelt dat een subjektief moment aanwezig is in de aanwijzing van het (meest) juiste alternatief in de skorings<i>sleutel</i>. M.a.w.: bij meerkeuze vragen doet zich exakt hetzelfde beoordelingsprobleem voor als bij kort-antwoordvragen. Zaalvoorzitter Hofstee

ontlokte dit de uitspraak dat hij het als een volstrekte eye-opener ervaren had. <p><br>

Wat het aantal alternatieven betreft wordt te vaak ervan uit gegaan dat dat vier zou moeten zijn; zelfs worden soms wanstaltige dubbelvragen bedacht om daaraan te kunnen voldoen. In 'Advies over aantal keuzealternatieven bij studietoetsvragen' (jan. 1976) wordt gedemonstreerd dat twee of drie alternatieven de technische kwaliteit van de toets tenminste niet zullen aantasten, en de problemen van het 'bedenken' van alternatieven sterk verlichten. Recent hebben Van de Brink en Van Naerssen hetzelfde uiteengezet in TOR 1979. <p><br>

Niet zelden wordt eerst gekozen voor het hanteren van de meerkeuze-vraagvorm, waarna alle vragen over de stof in dat (te) nauwe keurslijf worden gedwongen. Dat levert onderwijskundig bekeken nogal eens wanstaltige vragen op waar het beter geweest zou zijn om open eind of kort antwoord vragen te gebruiken, of essay vragen. Traditioneel worden deze verschillende vraagsoorten karakteristieke voor- en nadelen toegedicht die ze min of meer onafhankelijk van het soort leerstof en leerdoelen waar het om moet gaan, zouden hebben. Dat kan leiden tot ongelukkige a priori keuze van te gebruiken vraagsoort. In 'Toetsen' (1979) wordt het vraagstuk van de keuze van vraagsoort behandeld als een afgeleide van de aard van leerstof en -doelen: deze bepalen welke vraagsoort het best 'past'. Dat kan makkelijk leiden tot het gebruik van

verschillende vraagsoorten in één en dezelfde toets, en daar hoeft bij verstandig gebruik (goede layout, goed erop voorbereid zijn van de studenten, geen enkel bezwaar tegen te bestaan. Er zijn nogal wat situaties waarin zowel meerkeuze als kort antwoord vraagsoorten in dit opzicht goed bruikbaar zouden zijn: wanneer verondersteld mag worden dat bij de kort-antwoordvraag de studenten zullen kiezen uit een klein aantal hen goed bekende alternatieven. Zijn er echter studenten bij die deze alternatieven niet goed kennen, dan is de meerkeuze vraagsoort minder geschikt omdat dan de raadkans een rol gaat spelen. In dat verband realiseert men zich te weinig dat voor raadkansen niet gecorrigeerd kan worden in strikte zin, maar slechts op statistische wijze. Raadkansn zijn in dergelijke gevallen eenvoudig een vermijdbare extra bron van foutenvariantie, en zij vergroten ook de standaardmeetfout voor de gegeven student met ongunstige gevolgen voor de transparantie van de toetsing.

  

Toetsen bestaande uit meerkeuzevragen kunnen automatisch geskoord worden. Voor toetsen door de docent zelf gekonstrueerd is dat een twijfelachtig voordeel. Wordt de toets door 1400 studenten afgelegd (rechten in Leiden bijv.) dan weegt dat voordeel zwaar. Zijn er maar 100 tot 200 dan valt dat voordeel weg tegen de extra zorg die het konstreren van meerkeuzevragen veelal vraagt vergeleken met kort-antwoordvragen. De mogelijkheden om bij heel grote aantallen studenten hen het eigen werk te

laten skoren zijn nog onvoldoende onderzocht, of beter: nog nimmer serieus bekeken. In beginsel heb je bij dat laatste alleen een steekproefsgewijze controle nodig op alle deelnemers, en een volledige op alle deelnemers juist boven de cesuur (en degenen er juist beneden die misschien al te eerlijk zijn geweest en van zich af gerekend hebben).

<p><br>

Het bedenken van toetsvragen wordt door vrijwel alle docenten een heel lastige opgave gevonden. Voor meerkeuzevragen geldt zelfs dat het maken van dergelijke vragen als een 'kunst' beschouwd wordt. Moet je ieder jaar weer nieuwe vragen bedenken, en dat bovendien nog voor enkele herkansingstoetsen, dan is het duidelijk dat het tot nog toe ontbreken van heuristische voor het genereren van (onderwijskundig verantwoorde) toetsvragen een ernstig probleem vormt. Eind zestiger jaren werden oplossingen gesuggereerd (item forms, en het werk van Bormuth), die weinig levensvatbaar bleken. Op de intuïtie dat het werk van Gagne; (Conditions of learning) wel eens &acute;&acute;n van de sleutels zou kunnen bevatten heb ik in 1973 geprobeerd te doen wat Gagne; naliel: de consequenties van zijn eclecticische theorie voor het formuleren van vragen over de leerstof uitwerken. Een 2e bewerking kwam in 1976 tot stand, een 3e en 4e in 1979, de laatste als hoofdstuk 2 en 3 opgenomen in 'Toetsen' (1979). De opzet om heuristische te ontwikkelen die zowel eenvoudig hanteerbaar zijn, in korte tijd grotere aantallen

vragen opleveren, als onderwijskundig (d.i. vanuit de onderwijsdoelstellingen bekeken) zinvol zijn, is daarmee zeker geslaagd. Wat niet wil zeggen dat talrijke verbeteringen niet mogelijk zouden zijn. Dit werk toont verwantschap met dat van Klausmeier, en eveneens met een soort docentenkursus van Merrill (1976). <p><br>

Het begrip 'betrouwbaarheid' speelt bij toetsen een merkwaardige rol. Je vindt dat bijv. aangeduid door Stanley (1971) die stelt dat betrouwbaarheden alleen een rol spelen wanneer gekozen moet worden tussen het gebruik van ongeveer even valide toetsen, maar vervolgens niet schroomt over betrouwbaarheid een zeer lange monografie ten beste te geven. M&eacute;t Stanley, maar ook m&eacute;t Lord (1957) en vele anderen, ben ik van mening dat het begrip betrouwbaarheid bij het beoordelen in het onderwijs een volstrekt ondergeschikte plaats dient te krijgen, en met name ondergeschikt aan die van de standaard meetfout voor de gegeven persoon (zie wederom Lord 1957, en natuurlijk 'Toetsen' (1979)). De adstructie van deze opvatting is een veelzijdige, en werd in 'Toetsen' gegeven. Daarbij werd gewezen op de methodologische vergissing die vele auteurs onbevangen begaan: te spreken over betrouwbaarheden zonder daarbij aan te geven welke doelgroep van personen je daarbij op het oog hebt. (Dezelfde vergissing wanneer het gaat om predictieve validiteit wordt meestal expliciet niet gemaakt, daar is men zich meestal w&eacute;t;

bewust van fenomenen als 'restriction of range' waar rekening mee te houden is; welnu, hetzelfde moet je doen bij betrouwbaarheidsrekening). Opmerkelijk is dat in de generaliseerbaarheidstheorie van Cronbach c.s. (1972) de methodologie op dit punt perfect in orde is, maar dat de evidente doorverbinding naar het hanteren van het ouderwetse betrouwbaarheidsbegrip nog niet gemaakt is).<p>

Bij criterium-gerefereerde toetsing zou je verwachten dat er wat argwanender tegen die 'betrouwbaarheid' aangekeken zou worden, maar integendeel zien we daar juist een stortvloed van publikaties die pogen deze oude wijn in nieuwe zakken te gieten. (Vooral ook waar het gaat om 'betrouwbaarheid van zak-slaag beslissingen', wat flagrante flauwekul is).<p>

Waar het besef dat het veeleer gaat o de s.m. (standaarmetfout) voor de gegeven persoon dus alud is, is dat besef nimmer vertaald in een hanteerbare, noem het voor mijn part didakometrie. Een gedetailleerde eerste poging daartoe heb ik gegeven in hoofdstuk 8 van 'Toetsen' (1979), zij het dan ook dat heel wat evidente doortrekkingen daar nog niet gepleegd zijn (maar zelfs door kundige docenten z&eacute;lf uit te vinden zijn). Ik heb daarbij goed gebruik kunnen maken van statistische technieken die in de literatuur ver criterium-gereferrerde toetsing recentelijk tot wasdom gekomen zijn (zie ook de bijlagen bij 'Studiestrategi&euml;n' (1978)).<p><br>

De <i>interpretatie</i> (of analyse) van toetsresultaten

is in de literatuur altijd stiefmoederlijk behandeld: wat al te gemakzuchtig werden daar dezelfde technieken voor aangedragen als bij de psychologische testanalyse gebruikt werden. Inklusief natuurlijk de analyse op itemniveau, met zijn item-test correlaties, p-waarden, de hele psychometrische santekraam. Het zal de ervaring van velen geweest zijn daar in de praktijk niet zinvol mee uit de voeten te kunnen. In 'Toetsen' heb ik de kwaliteitskontrolle op items (dat zijn toetsvragen dus) weer zijn juiste plaats teruggegeven: de docent voert die kontrolle uit op basis van zijn specifieke deskundigheid als vakdocent, aangevuld met de inhoudelijk relevante oordelen van kollega-vakdocenten. Slechts in uitzonderlijke gevallen zullen itemstatistieken nog een ondermaats item boven water brengen. De interpretatie van de resultaten op de toets in zijn geheel, voor deze groep studenten, kan alleen zinvol zijn wanneer dat gebeurt in termen van de door deze groep bereikte beheersing van de stof, waarbij de laatste gedefinieerd is op een domein van vragen dat misschien re&euml;el bestaat , maar tenminste conceptueel omschreven is. Zie daarvoor hoofdstuk 8 van 'Toetsen (1979)). <p>Andere vormen van interpretatie, in de richting van verschillende vormen van validiteit, kunnen langs traditioneel bekende wegen worden gemaakt. Een goed hulpmiddel bij het ontwikkelen van vruchtbare vraagstellingen daarbij is het model voor studieprestaties, zoals samen met Tromp gepresenteerd op de ORD 1977: 'Het meten van studietijd'. Voor interpretaties m.b.t. studiestrategie&euml;n van studenten, en

cesuurbepaling, verwijs ik naar mijn studies op beide gebieden. <p><br>

Een algemene lijn in dit werk op toetsgebied is het verlaten van het onbereflekteerde uitgangspunt dat toetsen gelijk zou zijn aan 'meten', en daarmee ook als enigszins losstaand van het overige onderwijsgebeuren beschouwd zou kunnen worden, en het betrekken van de stelling dat de beoordeling een heel functionele plaats in dat onderwijsgebeuren heeft, of men dat nu expliciet wil onderkennen of niet. Natuurlijk is het beter om je af te vragen hoe de beoordeling in dat onderwijsgebeuren functioneert, dan is diezelfde beoordeling rationeel te hanteren als middel om datgebeuren (mede) te sturen. In 'Toetsen', zoals ook in 'Ceurbepaling' en 'Studiestrategie&uml;n', geef ik de inzichten en technieken die de docent opnieuw greep geven op zijn eigen onderwijs - en toetsingssituatie. <p><br>

TENTAMENMODELLEN (of: studiestrategie&uml;n).  
<P><BR>

Ik heb altijd de nadruk gelegd op afstemmen van het te geven onderwijs op de daaraan deelnemende studenten, i.t.t. selectie van studenten zodanig dat de overblijvenden goed bij het altijd al gegeven onderwijs passen (beide nota's 'Selectie contra onderwijs' in 1971 bijv.). Dat doet de vraag rijzen hoe je de beoordeling van studieprestaties in die

opvatting kunt inpassen. Het ligt voor de hand om dan aansluiting te zoeken bij de ideeën van Carroll en van Bloom, mastery learning dus, met als afgeleide daarvan criterium-gereferente toetsing. Maar in de kenmerkende benadering van de problemen rond criterium-gereferente toetsing is nog steeds de traditionele instelling voelbaar dat studenten zich ook moeten maar hebben te plooiën naar beoordelingsprocedures die meer institutioneel van karakter blijven dan dat zij gericht zijn op 'optimale' ontplooiing van de talenten van de studenten. [Echt waar, het staat er zo!, b.w. 2021]

Het gaat er om dat je enig inzicht krijgt in het samenspel, of tegenspel, van de wijze waarop de beoordeling is ingericht, en de wijze waarop studenten daarop reageren in hun studiestrategie, hun motivatie, of hun tijdbesteding. Zoals geformuleerd is het evident dat je allereerst iets over studiestrategieën van studenten zou moeten weten, om daarna zinvolle uitspraken te kunnen doen over beleidsalternatieven voor de docent (m.b.t. cesuurbepaling bijv.). Een eerste en belangrijke aanzet werd door Van Naerssen gegeven (1970). Ik heb geprobeerd onafhankelijk van Van Naerssen een (beter?) tentamenmodel te ontwerpen, maar kwam daarbij uit op een verbeterde versie van Van Naerssen's 1970 model (voor het geval van een conjunctieve regeling en onbeperkte herkansingsgelegenheid).

In een eerste reactie op 'Studiestrategie&uml;n' liet Van Naerssen mij weten dat hij blij was met deze overbrugging van de kloof tussen theorie en praktijk. M.a.w.: het tentamenmodel is nu van handen en voeten voorzien, en kan het veld in gestuurd worden.

<p><br>

Het gebruiken van het beta-binomiale model maakt een doorzichtiger en ook een preciezer aanpak mogelijk dan door Van Naerssen gegeven, en maakt het eenvoudiger om ook voor afwijkende onderwijssituaties modellen op te stellen (beperkte herkansingen, kompensatorische regelingen, en hybride propedeuse regelingen).

<p><br>

Ook maakt het beta-binomiale model een kwantificering van de cesuurbepalingsproblematiek, gebaseerd op tentamenmodellen, beter mogelijk.

<p><br>

Andere verbeteringen zijn: een vereenvoudigde inschatting van 'verliezen' bij het moeten herkansen; en het gebruik van een eenvoudige en hopelijk robuuste leercurve.

<p><br>

Uitwerking van dit tentamenmodel voor het geval van onbeperkte herkansingsgelegenheid leidde tot de ontdekking dat er een breed scala van ongeveer optimale studiestrategie&uml;n bestaat. Dat is vooral d&acute;&acute;rom een belangrijke vondst dat het gevoerde onderwijs-en toetsingsbeleid er toe zou kunnen leiden dat studenten stelselmatig strategie&uml;n kiezen die hen met een geringe

stofbeheersing laten slagen, waar alternatieve beleidsmogelijkheden hen er toe zouden kunnen brengen strategieën te volgen die bij eenzelfde verwachte tijdbeslag tot hogere stofbeheersing (en minder herkansingen) leiden. <p><br>

Is het aantal herkansingsgelegenheden per vak beperkt, dan ontstaat een geheel andere situatie, waarop het boven geschetste tentamenmodel niet meer van toepassing is. Nu is het voor de student allereerst van belang zijn zak-risico (inklusief herkansingen) te beheersen, waarvoor een model voorgesteld wordt. Uiteraard bestaat voor de laatste herkansingsgelegenheid de beste strategie er meestal uit om zo hard mogelijk en tot het laatste moment te werken, waarmee een meestal niet gewenst gevolg van het beperken van herkansingen aangetoond is. (Alternatieven zijn wel beschikbaar, en o.a. hfdst 9 van 'Studiestrategieën ter sprake gebracht'). <p><br>

Ook werd een tentamenmodel voor kompensatorische regelingen ontwikkeld, met als bijzonderheid daarbij de 'wisselstrategie': het is voor de student mogelijk om op een enigszins riskante studiestrategie te werken, zolang er voldoende ruimte blijft om zonodig over te schakelen op een hogere strategie die het risico aanvaardbaar maakt. ('Studiestrategieën' par. 8.4). Dit tentamenmodel heeft eveneens handen en voeten in deze zin dat er een duidelijk beleid door docent (en/of studieleiding) op te baseren is. <p><br>

Een aantal gebruikelijke propedeuse regelingen die tussenvormen van conjunctief en kompensatorisch toetsen behelzen, worden eveneens kort gemodelleerd en op hun consequenties geanalyseerd (Studiestrategieën; hfdst 9). <p><br>

Afzonderlijk is aandacht besteed aan verschillende vormen van leercurven die onder uiteenlopende omstandigheden verwacht kunnen worden, en hoe deze tot dezelfde grondvorm terug te brengen ('Studiestrategieën' bijlage D). <p><br>

Wat is de betekenis van al deze modellen? In de eerste plaats zijn ze door mij ter gelegenheid van de docentenkursus 'Studiestrategieën' *normatief* bedoeld: studenten zouden op basis van dit werk 'rationele' studiestrategieën *bewust* kunnen kiezen, terwijl docenten hen daarbij zouden kunnen steunen door relevante informatie, kwantitatieve gegevens, tabellen, proeftentamens, te verstrekken. In de tweede plaats zijn deze modellen van belang als een goede start voor een *beschrijvende* theorie van hoe studenten studeren, als zodanig in de praktijk onderzoekbaar, zo je wilt ook falsificeerbaar. De laatste gedachte wordt door mij nog niet volledig omhelsd, ik zal daar o.a. in Cesuurbepaling concreter vorm aan geven omdat juist *de empirie* direkt aan de orde is.

Ten derde suggereren deze modellen talrijke verbanden tussen beleid en mogelijke gevolgen die geenszins triviaal of vanzelfsprekend zijn; in die zin vormen ze een verrijking van het arsenaal aan hypothesen waarover de adviserende onderwijskundige kan beschikken. <p><br>

Het belang van dit werk werd door Hofstee in zijn voordracht 'Toekomstige onderwijswetenschap' t.g.v. de oprichtingsvergadering van de Werkgemeenschap Research Hoger Onderwijs (18-5-79) aangegeven: <ul><li><i>"Nadruk op onderwijskundige modellen. Een uitstekend voorbeeld vormen de tentamenmodellen van Van Naerssen, naar mijn mening veruit het belangrijkste onderwijswetenschappelijke werk dat in Nederland is verricht. Een verwante ontwikkeling is het werk van Wilbrink m.n. over compensatorische examenregelingen, en dat van De Gruijter en anderen over caesurbepaling." "Maatregelen die uit traditie of uit normatieve a priori's worden gehanteerd kritisch op hun feitelijke werking onderzoeken, bijv. toetsnormen en examenprocedures bezien op hun effectiviteit om het leergedrag te beïnvloeden (wederom Wilbrink)." </i></li></ul>Publikatie van deze tentamenmodellen in het TOR ligt voor de hand, maar is wegens andere activiteiten (Toetsen, Cesurbepaling, criteriumgerefereerde grensscores naar de nabije toekomst (in de loop van dit jaar) verschoven. <p><br>

## TOELATING TOT NUMERUS FIXUS STUDIES <p><br>

De problematiek van de wijze van toelating tot numerus fixus studierichtingen is nauw verwant aan die van selectie voor het wo in het algemeen, en aan die van de selectieve propedeuse. Beide laatste waren, eind zestiger, begin zeventiger jaren in discussie (de nota's van Posthumus, ontwerp van wet herstructurering w.o. 1971, De Groot 'Selectie voor en in het hoger onderwijs' 1972, de RWO brochure 'Herstructurering zonder herprogrammering' 1972).  
<p><br>

In de discussienota 'Selectie contra onderwijs' (met Wybe Zijlstra, Subfakulteit scheikunde, maart 1971) wordt een pleidooi gehouden om van selectieve beoordeling tot evaluatieve beoordeling te komen. In het gelijk getitelde 'Selectie contra onderwijs' (Wilbrink, september 1971) is de analyse verdiept, en wordt de traditionele vraag 'heeft deze student het onderwijs met vrucht gevolgd?' gekontrasteerd met het alternatief 'hoe geven we een vorm aan ons onderwijs die de student in staat stelt het onderwijs met vrucht te volgen?' In beide stukken wordt de selectie als panacee voor de problemen in het w.o. gerelativeerd; de gegeven visie op de plaats die het beoordelen in het onderwijs zou moeten hebben, laat al de belangrijke elementen zien die in 'Toetsen' (1979) en in het werk over tentamenmodellen (1978) uitgewerkt zijn.  
<p><br>

Een wel heel bijzondere wijze waarop het belang van selectie overschat wordt ten koste van aandacht voor een meer leerling-vriendelijke inrichting van het onderwijs blijkt uit het streven om kompensatorisch onderwijs te richten op verbetering van vaardigheden (intellectuele) die van belang zijn bij het goed kunnen maken van tests (intelligentie tests) die bij selectie in het onderwijsstelsel een rol spelen. Dat verleidde velen tot tendentieuze en averechtse interpretaties van Jensen's (1969) artikel 'How much can we boost I.Q. and scholastic achievement?', hier te lande door De Groot in discussie gebracht. Ik heb er in 'Spiegelgevechten om het I.Q.' (27-9-72) op gewezen dat het Jensen er juist om te doen was aan te duiden dat het uitblijven van succes van kompensatorisch onderwijs alles te maken heeft met het gebruiken van een verkeerde maat daarvoor (verbetering van intelligentietest scores), en wegen aan te geven waarin datzelfde kompensatorisch onderwijs verbeterd kan worden. Ik wees er daarbij op dat, wat genetische bepaaldheid van verschillen betreft, het juist is dat succesrijk kompensatorisch onderwijs er toe zal leiden dat de proportie van daarna nog overblijvende verschillen hogere mate genetisch bepaald zal zijn omdat immers aan het milieu toe te schrijven verschillen gereduceerd zijn. <p><br>

November 1971 wordt een gedetailleerde studie 'Levert selectie in het hoger onderwijs iets op?' afgerond. Het aksent ligt daarin op 'predictieve validiteit' en 'rendement van de selectie', waarbij een

besliskundige benadering gehanteerd werd. Selectie in de propedeuse, of een extra selectie voorafgaand aan de studie, kunnen in de bestaande Nederlandse onderwijsstructuur (een sterk selektief middelbaar onderwijs) geen nut afwerpen. Doorrekenen van een model van Van Naerssen (1970) zou ertoe leiden dat in een optimale selectieprocedure ongeveer 2 % uitgeselekteerd zou mogen moeten worden, en dat lijkt de grote investeringen en de nadelige gevolgen van extra selectie niet waard. Afwijzen van grotere aantallen kandidaten zou willekeurig zijn, en de afgewezenen een veelal onterecht odium van 'mislukt' te zijn meegeven. In de konklusies wordt dan ook sterk gesuggereerd een eventueel noodzakelijke getalsmatige beperking (numerus fixus, naar later zou blijken) via een <i>lotings</i>procedure te bewerkstelligen. Dat is een faire methode naar de kandidaten toe (in afwezigheid van voldoende predictieve validiteit van andere gegevens), belast afgewezenen niet met de frustratie het 'niet gekund te hebben', heeft geen nadelige effecten op het onderwijs. <p>

Weinig kon ik toen vermoeden hoe heet dit hangijzer in de komende jaren, nadat De Braauw met een lotingsvoorstel naar de Kamer kwam, zou worden!  
<p><br>

Maar eerst zou de studie van De Groot over selectie voor en in het hoger onderwijs nog verschijnen, waarin tal van misvattingen nog eens uitvoerig geëml;taled werden. Het COWO reageerde daarop bij de pen van Kolthoff met 'De gekooide

vogel' en 'Een duik in de valkuil' in de HP; de gelijktijdige publikatie van Hofstee's kritiek in De Groene was frappant unisono tot in technische details. Op verzoek van de Overlegcie. Tertiair Onderwijs (OTO) heb ik De Groot's studie op tegenspraken uitgelopen: 'Tegenspraken typeren het selectie-voorstel De Groot'. Wederom een studie naar verborgen vooronderstellingen bij voorstanders van (extra) selectie. <p><br>

Het verschijnen van de aanbevelingen van de werkgroep-Wiegersma aan de toenmalige minister De Braauw over toelating tot numerus fixusstudies (vnl. de strijd om loten tot 7 of tot 7,5), noopte CRWO tot het innemen van een duidelijk standpunt, uitgewerkt door Wilbrink en Van der Vleugel: 'Bij beperkte toelating beslist Lot' (OvO april 1974). De argumenten in dat artikel zijn een uitwerking van wat al in 1971 in 'Levert selectie in het hoger onderwijs iets op?' aangegeven werd. Een felle campagne van ene hoogleraar Bakker uit Leiden in de NRC/Handelsblad noopte tot een vervolgartikel in OvO wederom met Van der Vleugel 'Bij gebrek aan beter', en tot een artikel in de NRC/H. 'Selectie voor numerus is irrationeel'. <p><br>

In 'Bij gebrek aan beter' wordt het onderscheid tussen selectie op 'geschiktheid' en 'quota selectie' benadrukt. Omdat argumenten m.b.t. 'rendement' in de discussie een grote plaats innemen, werden enkele berekeningen gegeven om de werking van verschillende toelatingsprocedures t.o.v. elkaar te

kwantificeren. Voor de eerste keer wijs ik er op dat niet vergeten mag worden dat afgewezen kandidaten een andere studie, opleiding, of beroep zullen kiezen. *"Beslissingen bij numerus fixus-regelingen worden daarmee in belangrijke mate een kwestie van <b>plaatsing</b>, niet van <b>selectie</b> (zoals impliciet door de werkgroep-selectie verondersteld werd). Konkreet betekent dit dat niet aangenomen mag worden dat het rendement van afgewezen 'geschikte' kandidaten veel kleiner zal zijn dan het rendement van toegelaten geschikten; waarbij rendement betekent rendement voor het w.o. of rendement voor de maatschappij als geheel i.p.v. rendement voor de gelimiteerde fakulteit."* Een cijfermatige illustratie werd tevens gegeven. Vier jaar later zou het nog steeds nodig blijken geduldig uit te leggen dat bij numerus fixus-toelating niet zozeer de predictieve validiteit van eindexamencijfers (toelatingstoets), als wel de <i>differentiële</i> validiteit van belang is, en dat je voor dit soort gegevens mag verwachten dat de laatste zeer gering zal zijn. <p><br>

Vanuit iets andere invalshoek heb ik in 'Selectie voor numerus fixus is irrationeel' eveneens het <i>plaatsings</i> i.p.v. selectie karakter van deze toelatingsproblematiek uiteengezet. In dit artikel werd bovendien geschetst hoe verschillend de selectieproblematiek (voor de n-f studie) er uit ziet voor enerzijds de betrokken kandidaten, anderzijds de selekterende instanties.<p><br>

Een eerste besliskundige onderbouwing van deze verschillende toelatingsmodellen voor de diverse betrokkenen werd april 1975 gepresenteerd in 'Overzicht van de selectieproblematiek bij numerus fixus regelingen'. De pay-off van dit interne rapport ligt echter vooral in mijn latere studies m.b.t. criterium-gerefereerde grensscores. <p><br>

Een niet alleen technisch-statistisch interessant interpretatieprobleem deed zich voor bij de gewogen loting zoals die uiteindelijk bij amendement-Vermaat door de Tweede Kamer vastgelegd werd. In een technische studie, die tevens het functioneren van deze gewogen loting onder verschillende (quota-)omstandighederi demonstreerde, zijn beide alternatieve interpretaties uitgewerkt en geadstrueerd ('Gewogen loting', mei 1975).<p><br>

In de hierop volgende, althans wat 'loten' betreft wat rustiger tijd werd 'Selectief literatuur overzicht m.b.t. beoordeling en selectie, i.h.b. bij de propedeutische studiefase' samengesteld. <p><br>

Dan wordt in juni 1978 voor het CvB 'Reactie op het rapport van de werkgroep-Wiegersma' geschreven. Met de ondertussen bekende argumenten, en voorzien van een modelmatige illustratie van de werking van zo'n voorgestelde 'toelatingstoets'. Op basis van vnl. ditzelfde stuk wordt ook het CRWO standpunt bepaald. <p><br>

De publikatie van het CITO-rapport over de

mogelijkheden van een toelatingstoets noopte tenslotte tot het schrijven van een uiteenzetting over wederom de misvatting als zouden predictieve validiteit, en niet differentiële validiteit van zo'n toets van belang zijn: 'Toelatingstoets voor het wetenschappelijk onderwijs?' (1980, TOR).

Hiermee komt dan een (voorlopig) einde aan een lijn van theoretische studies die voornamelijk door actuele gebeurtenissen werd gestuurd, wat niet wegneemt dat de bereikte inzichten hun rol gespeeld hebben bij studies over andere beoordelingsproblemen (in het w.o.).

PM Een bladzijde over het bibliotheekbeheer dat mij in 1968 was opgedragen.

PM Lijst van schriftelijke stukken, 4 blz. Steekwoordsgewijs: 68 MDF Eindhoven, Overleg en zeggenschap, discussinota Scheikunde, selectie Helvoirt (ook als advies aan UR), Soiegelgevechten HP, Tegenspraken OTO Doelstellingen Thk, Beoordelingsmethoden FSW-b doelstellingen rechten, enz. Ik check de lijst nog aan mijn publicaties.htm.