

[figuren ontbreken in de OCR-file; kan ik deze reconstrueren?]

overdoen afschaffen

18-8-76

Methoden voor het bepalen van de grens zakken slagen voor studieonderdelen. Of: een aantal onafhankelijke argumenten voor het afschaffen van de gewoonte om onvoldoendes te laten inhalen.

Concept voor docentencursus moduul. Ben Wilbrink.
18-8-76.

Voorwoord.

Dit rapport is inderdaad een beetje tweeslachtig: ik wil er een schets in geven van de aanpak van het probleem hoe je zak-slaagbeslissingen voor afsluitende tentamens op verantwoorde wijze kunt nemen. Hoe je dat probleem ook aanpakt, de oplossing leidt in ieder geval altijd tot het afschaffen van een merkwaardige gewoonte die wij Nederlanders (en vele andere Europeanen) in ons onderwijs koesteren: nogal wat onvoldoendes uitdelen, en daarbij ook nog eisen dat de slachtoffers de betreffende studieonderdelen overdoen tot ze wél voldoende zijn. Vandaar dat ik naast beschrijving van de aanpak van het cesuurbepalingsprobleem, ook een reeks argumenten presenteer om in ieder geval het instituut 'onvoldoendes worden ingehaald' af te schaffen. De argumenten voor afschaffing van het overdoen, zoals ik het voortaan zal noemen zijn nogal uiteenlopend van aard, voor de lezer waarschijnlijk in hoge mate verrassend, en naar mijn bescheiden mening ieder voor zich al voldoende om tot afschaffing van het overdoen over te gaan.

Het is niet onwaarschijnlijk dat bij verder lezen in dit rapport de lezer geneigd is om de kwalificatie 'verrassend' als nogal eufemistisch te gaan beschouwen. Veel argumentaties zullen op het eerste gezicht onwaarschijnlijk, onjuist, of weerzinwekkend lijken. Daarom wil ik hier in het kort ingaan op aard en functie van onderwijskundige

producties als deze.

De docent is, niet onbegrijpelijk, geneigd om de boodschap van onderwijskundigen te toetsen aan zijn eigen ervaringen in het onderwijs, zowel waar het om het geven van onderwijs gaat, als om inzichten in capaciteiten van studenten, de beste methode om kennis en inzicht van die student te toetsen, en de wijze waarop je selectie voor het onderwijs zou moeten inrichten. De inzichten die de docent zich al doende eigen maakt omtrent onderwijs en alles wat er zoal mee samenhangt, zijn echter noodzakelijk onsystematisch, gekleurd door persoonlijke opvattingen en toevallige ervaringen, beïnvloed door niet noodzakelijk juiste 'common sense' ideeën. De gedachtewereld van de docent staat bijna per definitie op gespannen voet met die van de onderwijskundige. Daarmee bedoel ik het volgende: het is eenvoudig niet de taak van de onderwijskundige om de docent te bevestigen in zijn ideeën over hoe je onderwijs verzorgt. De samenleving heeft geen behoefte aan onderwijskundigen die bestaande onderwijs praktijken bevestigen (tenzij toevallig het resultaat van alle onderzoek zou zijn dat de manier waarop we het tot nu toe gedaan hebben, inderdaad 'de beste' is), maar aan onderwijskundigen die direct of indirect de nodige informatie verschaffen om dat onderwijs te verbeteren (vernieuwen).

Dezelfde gedachtegang is van toepassing op de vraag op welke punten dat onderwijs dan wel in eerste instantie voor verbetering vatbaar zou zijn: ook daar is vaak te constateren dat er een grote afstand bestaat tussen wat de docent, in de praktijk van het onderwijs, als de belangrijkste tekortkomingen van dat onderwijs ziet, en wat de onderwijskundige aanwijst als de belangrijkste te repareren tekorten. De kloof in denkwereld is in dit geval ernstiger dan het verschil van mening dat je met de loodgieter kunt hebben over wat de oorzaak van een bepaalde lekkage kan zijn: we zijn nu eenmaal makkelijker geneigd een vraagteken te plaatsen achter onze amateuristische kennis van het 'loodgieten' dan achter onze 'kennis van het onderwijs'. En ach, dat geeft ook niet, zolang beide partijen zich er maar een beetje bewust van zijn: de onderwijskundige dat hij voor zijn uiteenzettingen niet direct op een al te willig gehoor mag rekenen, en de man of vrouw uit de praktijk dat hij er niet op moet rekenen van die onderwijskundige argumenten te horen die hem of haar direct zullen bevallen.

Zo ook met dit rapport. Het zou niet geschreven zijn, wanneer de inhoud ervan niet verrassend zou zijn, wanneer de docent zich af en toe niet eens behoorlijk aan de 'vreemdheid' van de hier gepresenteerde ideeën zou stoten. Met mijn waarschuwing hoop ik te bereiken dat het rapport wat minder in ergernis, en wat meer bereidheid om althans de argumenten een kans te geven, gelezen wordt. Dan zullen er toch nog punten van commentaar en kritiek genoeg overblijven: wat dat betreft ben ik bereid op alle opmerkingen te antwoorden. Maar..., ik wil daar direct het volgende aan toe voegen:

De lezer die overloopt van 'kritische kanttekeningen' moet zich niet al te veel illusies maken over zijn 'originaliteit' daarbij: de meeste bedenkingen komen voort uit een onvoldoende kennis van zaken. Ik geef onmiddellijk toe dat de presentatie van argumenten in dit rapport de nodige weerwoorden uitlokt, maar de kritische lezer mag bedenken dat vrijwel iedere uitspraak gebaseerd is op relevante literatuur waarmee hij weinig of in het geheel niet vertrouwd is. Vertrouwt hij die literatuur (zowel theoretisch van karakter als empirisch materiaal, bevattend) niet, dan moet ik hem (haar) aanraden met mij contact op te nemen: ik heb het voornemen om een 'schaduw rapport' uit te brengen, bedoeld voor onderwijskundige collega's in het land, dat beknopter van tekst zal zijn, maar waarin uitvoerig beschreven zal worden op welke bronnen uitspraken en ideeën gebaseerd zijn.

Daarnaast geldt, wat uitwerking van de methode voor cesuurbepaling betreft, dat de praktische uitvoerbaarheid van die methode in dit rapport nauwelijks geïllustreerd kan worden; aan het laatste is een afzonderlijk rapport te wijden, waarin de Ischroefdes en moertjes, van de methode besproken zullen worden. De bedoeling van het nu voor u liggende rapport is wat de methode van cesuurbepaling betreft, een heel bescheidene: de docent overtuigen van de mogelijke zinvolheid van deze methode, om hem of haar te motiveren het beloofde technische rapport te bestuderen op het punt van toepasbaarheid voor zijn of haar onderwijssituatie.

Een aantal lezers zullen kritiek hebben op onderdelen van het rapport die hun eigen vakgebied raken. Voor die kritiek

geldt, uiteraard ook dat ik er bijzonder in geïnteresseerd ben. (Ik denk aan filosofische, psychometrische, besliskundige, economische, didactische, differentiële psychologische, en andere problemen die direct of zijdelings door mij aangeroerd worden).

Als dit voorwoord als bijzonder pedant is overgekomen, spijt mij dat vreselijk: het was zeker niet mijn bedoeling. De onderwerpen die ik ga aansnijden zijn echter nogal controversiële van aard, en mijn ervaring heeft mij geleerd dat bijzondere maatregelen dan niet overbodig zijn.

Voorschouw van de argumenten voor afschaffing van het overdoen.

De argumenten voor het afschaffen van overdoen zijn verschillende, bovendien van elkaar onafhankelijke argumenten. Dat wil zeggen dat het mogelijk niet geheel overtuigend zijn van een enkele wijze van argumenteren de kracht van de andere argumenten onverlet laat. Of, anders gezegd, de waarde van de ene argumentatie is er niet van afhankelijk of de lezer een andere argumentatie al dan niet overtuigend vindt. De volgende opsomming is dan ook echt een opsomming en niet een betoog op zich.

Argument 1. Het zal bijzonder moeilijk aantoonbaar zijn dat studenten die ondanks behoorlijke voorbereiding een onvoldoende scoorden, bij herhaling van het tentamen de stof werkelijk beter beheersen. Zorg er voor dat alleen studenten die zich behoorlijk voorbereid hebben aan tentamens deelnemen, en het laten overdoen van onvoldoendes kan achterwege blijven (of: en er hoeven vrijwel geen onvoldoendes meer te vallen).

Argument 2. Voor studenten die ondanks behoorlijke voorbereiding een onvoldoende scoorden, zal een effect op verder studiesucces van het 'verbeterd' hebben van die voldoende, in gunstige zin niet aantoonbaar zijn. Dit soort overdoen dient geen enkele doelstelling van onderwijs, en kan beter achterwege blijven.

Argument 3. Het laten overdoen van onvoldoende gemaakt

studieonderdelen veronderstelt impliciet dat het laten overdoen een betere procedure is dan het gewoon door laten gaan met de studie. Die aanname is in vrijwel alle onderwijs situaties aantoonbaar onjuist, waarmee tevens de irrationaliteit van het laten overdoen aangetoond is.

Argument 4. Aan het laten overdoen van studieonderdelen zitten vrijwel altijd selectieve bedoelingen vast. Het is de taak van de docent om vast te stellen wat een voldoende, en wat een onvoldoende prestatie voor zijn vak is. Maar het is niet de taak van de docent om de student onder druk te zetten en de studie maar te staken: áls er geselecteerd moet worden, is dat een taak voor de subfaculteit. Welnu, wanneer de docent zich bij zijn taak houdt', en doceert in plaats van selecteert, is er voor het instituut 'overdoen' geen plaats meer.

Argument 5. De subfaculteit die meent dat zij goede argumenten heeft om studenten nog te selecteren nadat zij zijn toegelaten tot het onderwijs, staan daartoe veel beter beleidsmaatregelen ter beschikking dan selectie over te laten aan de chaotische uitwerking van de opeenvolging van zak-slaagbeslissingen door onafhankelijke docenten genomen. Een voorbeeld van een beoordelings systeem waarin selectieve beslissingen genomen worden zonder dat studenten eindeloos veel tijd aan het onerdoen van studieonderdelen hoeven te verliezen is het Amerikaanse Grade Point Average systeem (zie Bevers 1975).

Argument 6. Afschaffen van het overdoen zou de gemiddelde studieduur voor degenen die afstuderen met tenminste een half jaar bekorten. Het economisch voordeel daarvan is te schatten op een bedrag van enkele honderden miljoenen guldens jaarlijks.

(Anders 'gezegd: de nodeloze verstrengeling van cesuurbepaling en selectie brengt hoge kosten met zich mee: een groot aantal ten onrechte uitgedeelde onvoldoendes en daaruit voortvloeiende studievertraging, hogere onderwijslasten, e.d.)

Argument 7. In het stelsel van 'overdoen' zijn de prestatie eisen dermate hoog opgeschroefd dat de daarvoor benodigde extra

studietijd(ook al door de extra tijd die verloren gaat doordat overdoen interfereert met de gewone studieprogrammering) in geen enkele reële verhouding meer kan staan met de eventuele leerwinst (ook voor degenen die wél in één keer een voldoende behalen).

Argument 8. Gesuurbepaling en overdoen zoals nu gebruikelijk, zijn in volstrekte tegenspraak met de meest minimale normen die aan toetsing in het onderwijs, en beslissingen over studenten in dat onderwijs, gesteld moeten worden. (Normen, zoals vastgelegd in de Standards for educational and psychological tests, van de American Psychological Association, American Educational Research Association, en National Council on Measurement in Education.)

Argument 9. In de meeste gevallen is toetsing op grond waarvan onvoldoendes worden uitgedeeld, onderwijs vijandig, in deze zin dat studenten de beslissing te weinig kunnen beïnvloeden door het volgen van een efficiënte studiestrategie. (Door onvoldoende transparantie van de beslissings situatie),

Argument 10. Het stelsel van over laten doen past in een opvatting over onderwijs die langzamerhand toch wel als achterhaald mag worden beschouwd: dat de studenten er zouden zijn voor het onderwijs, en niet het onderwijs voor de studenten verzorgd zou moeten worden.

Argument 11. Voor de huidige aanvechtbare procedures waarop uiteindelijk zak-slaagbeslissingen gebaseerd worden, zijn zonder veel moeite een groot aantal betere alternatieven te formuleren* (Die alle gaan in de richting van ofwel grotere transparantie, ofwel dat de verbetering van het onderwijs meer vruchten afwerpt dan oude beoordelingsgewoonteng of beide).

Argument 12. De zakslaag grens pleegt bepaald te worden zonder ooit daarbij expliciet de consequenties van daaruit voortvloeiende beslissingen over studenten te overwegen. Dat is een irrationele werkwijze die personen, én het onderwijsj grote schade berokkent. Technieken voor een meer met de rationaliteit op bevriende voet staande procedure, zijn al langé tijd in de literatuur bekend

(besliskundige aanpak).

Een alternatief voorstel voor de cesuurbepaling (de grote lijnen).

Het cesuur probleem als PROBLEEM schijnt eigenlijk alleen in Nederland enige aandacht te hebben gekregen. Enige uitzondering daarop zouden dan enkele publicaties zijn die zich bezig houden met het iets andere probleem. van het cijfergeven zoals dat in het Amerikaanse Grade Point Average systeem bestaat (Nedelsky, Downie, Ebel). In Nederland was Posthumus één van de eersten die de eigenaardigheden van onze onderwijs beoordelingsgewoonten bloot legde. Deze vijven-en-zessen-problematiek werd door De Groot uitgebreid beschreven in zijn bekende **Vijven en zessen**. Vlak vóór publikatie van Vijven en zessen presenteerde De Groot in Paedagogische Studiën een methode voor het bepalen van de cesuur: de kernitern methode.

In vrijwel ongewijzigde vorm werd deze methode overgenomen in Studietoetsen (1969). Van Naerssen werkte een modificatie van de methode uit, die eigenlijk van de uitgangspunten van De Groot weinig meer overliet, maar toch nog als kernitemmethode aangeduid werd, en als vervanging van de oudere methode opgenomen werd in de nieuwe editie van Studietoetsen (1975). De kernitemmethode heeft zich niet in discussie over haar verdiensten of onverdiensten mogen verheugen.

Uitzondering daarop is een korte bespreking in kritische termen door Wijnen, die tevens een alternatief presenteerde, bekend geworden als de 'methode Wijnen' (Swets en Zeitlinger? 1972). Grote verdienste van de bijdrage van Wijnen is . dat er een grondige correctie mee gegeven werd op de misstand die uit hantering van de kernitemmethode voortvloeide: de veelal, fantastisch hoge afwijzingspercentages werden in de methode Wijnen tot meer aanvaardbare proporties teruggebracht.

Over de beide kernitemmethoden (van De Groot en van Van Naerssen), en de methode Wijnen zullen we in het vervolg nog uitgebreid in de gelegenheid zijn kritische analyses te maken. Voorlopig is het voldoende op te merken dat in al deze pogingen voor het aanreiken van een cesuurbepalingsmethodiek verschillende heren

tegelijk gediend moesten worden: beoordeling, cesuur, en selectie werden niet van elkaar onderscheiden; cesuurbepaling bleef impliciet een middel waarmee tevens selectieve bedoelingen gerealiseerd moesten worden. Géén van deze methoden, hoewel op grote schaal in de praktijk toegepast (vooral de kernitemmethoden), is dan ook bevredigend gebleken, niet in de laatste plaats door het ontbreken van een, behoorlijke argumentatie voor juist déze methode. Dat geldt ook voor de methode Wijnen, die een aantal willekeurige momenten bevat die de docent voortdurend het gevoel geven op glad ijs te staan: de keuze van het gemiddelde als referentie punt, en twee keer de standaard meetfout als correctie daarop (zie Wijnen 1972). Hoewel Wijnen bij zijn zoeken naar een goede methode er expliciet van uitging dat een directe argumentatie voor een bepaalde cesuur vrijwel niet gegeven kan worden omdat daarvoor de nodige gegevens over de gehele onderwijssituatie ontbreken, waarbij hij uitkwam op een methode die volstrekt relatief zou moeten zijn, kon ook hij er niet onder uit om die relatieve methode aan enkele nogal willekeurige ankers op te hangen.

Een bevredigende aanpak van het cesuurprobleem lijkt dan ook slechts mogelijk door de zaak grondiger aan te pakken, met name ook door er niet van uit te gaan dat bestaande beoordelingsgewoonten gehandhaafd moeten kunnen worden. Het cesuurprobleem hoeft in het geheel geen probleem van 'vijven en zessen' te zijn. En inderdaad, in een ander bekend beoordelingssysteem, het Grade Point Average systeem, bestaat geen vijven en zessen problematiek.

Schets van de cesuurbepaling, zoals door de vakgroep uit te voeren.

Uitgangspunt is dat studenten die aan het studieonderdeel deelnemen, en de voor dit studieonderdeel uitgetrokken tijd er ook aan spenderen, binnen, die tijd een voldoende prestatie kunnen leveren. De moeilijkheid is hier de afstemming van studietijdbesteding op wat als minimaal aanvaardbare studieprestatie zal tellen. Behalve in practicum. situaties en dergelijke is immers de tijdbesteding van de student niet erg doorzichtig, er met name in de regionen waar de

slechtere studieprestaties geleverd worden is het niet eenvoudig uit te maken welke prestaties slecht zijn omdat de student zich onvoldoende heeft voorbereid, en welke prestaties slecht zijn omdat de student ondanks een (in tijdbeslag) voldoende voorbereiding niet tot betere resultaten kan komen.

Er lijken op voorhand twee mogelijke wegen te zijn om de situatie in de hand te houden. De meest directe is om de tijdbesteding van de student ook te meten (methoden voor studietijd meting zijn in principe wel beschikbaar) en de beoordeling voldoende/onvoldoende af te laten hangen van de uitkomsten van zowel eindtoets als tijdmeting. Keuze van het uitgangspunt genoemd aan het begin van de paragraaf betekent echter dat vooral gewicht gegeven zal worden aan de resultaten van de tijdmeting. Welnu, in deze omstandigheden worden aan de tijdmeting hoge eisen gesteld wil zij betrouwbare resultaten opleveren: bruikbaarheid van de methoden hangt af van de oprechte medewerking van de studenten. Waarschijnlijk mag op die oprechte medewerking meestal wel gerekend worden, zeker waar juist déze cesuurbepaling veel minder verstrekkende gevolgen voor de student heeft dan de nu gebruikelijke werkwijzen. Ook dan blijft nog als bezwaar bestaan dat de vrij veel invultijd vergende tijdschrijfmethoden als in wezen vreemd element in het onderwijs binnen gehaald worden. In dat licht bezien zou de techniek slechts bruikbaar zijn om af en toe, en dan nog steekproefgewijs, aanvullende informatie over het functioneren van het onderwijs te verzamelen.

Een meer indirecte weg is om de minimum eisen te formuleren in inhoudelijke termen, en door het creëren van zo transparant mogelijke toetsingssituaties er voor te zorgen dat ook de zwakkere studenten in de gelegenheid zijn aan die inhoudelijke minimum eisen te voldoen. Het opstellen van die minimum eisen is een proces dat gepaard zal gaan met herhaalde aanpassingen op grond van informatie over wat in de praktijk van dit stuk onderwijs haalbaar blijkt te zijn. Het is niet te verwachten dat de docent in één keer een pakket minimum eisen kan formuleren, waaraan vrijwel alle studenten ook binnen de daarvoor in het onderwijsprogramma uitgetrokken tijd, kunnen voldoen. Speciale methoden en technieken moeten dan

gebruikt worden om ofwel een controle te hebben op de tijdbesteding van de student (om de cesuurbepaling op deze wijze terecht te kunnen uitvoeren, en ' niet om die student te Big Brotheren), ofwel de wijze van beoordelen en de inrichting van het onderwijs een dusdanige vorm te geven dat daardoor alleen al in hoge mate een intensieve tijdbesteding van de student gewaarborgd wordt. Daarnaast zijn nog kleine onderzoekjes mogelijk om na te gaan of men met de cesuur zoals nu inhoudelijk vastgelegd, op de juiste golflengte zit.

Zoals hier uiteengezet, hangt de formulering van een minimum eisen pakket sterk af van de mogelijkheden om de doelstellingen voor dit studieonderdeel in inhoudelijke betekenisvolle termen te formuleren. De niet direct aanlokkelijke omweg van het eerst moeten formuleren van de doelstellingen van dit stuk onderwijs, zou bij een eerste toepassing van deze cesuurmethode vermeden kunnen worden door de minimumeisen vast te leggen in een opsomming van toetsvragen die tot op omschreven hoogte tenminste goed gemaakt kunnen worden. Dat kan een vragenlijst zijn waaruit ook de vragen voor de eind toets gekozen worden. Dan kan echter ook een opsomming zijn van ' de soort vragen die de student op de eindtoets kan verwachten, en die hij totop gespecificeerde hoogte goed moet kunnen maken. In het laatste geval is het belangrijk er zorg voor te dragen dat de omschrijving van het soort vragen de student ook voldoende duidelijk is, en ondersteund wordt door eer. voldoende aantal voorbeelden van de bedoelde vragen.

Op de wát langere termijn zal toch overgegaan worden op een wat minder intuïtieve wijze var. inventariseren van de minimum doelstellingen. De wijze waarop doelstellingen in inhoudelijke termen zinvol formuleerbaar zijn, bovendien in een vorm die zich leent tot uitgangspunt voor het bijzonder efficiënt en uitgebreid kunnen formuleren van evident zinvolle vragen om in de toetsing op te nemen, zal in het volgende, meer praktische, rapport over de concrete uitvoering van deze cesuurmethode aan de orde komen.

Het bovenstaande is in tamelijk abstracte termen geformuleerd, en met name is daarin niet gerefereerd aan de praktische moeilijkheden

die voortvloeien uit de onnauwkeurigheid van meten die inherent is aan toetsing van studieprestaties, en overigens aan alle metingen in de psychologie en de sociale wetenschappen. Dat betekent dat onze misschien zo aardig geformuleerde minimale eisen getoetst moeten gaan worden met sterk feilbare hulpmiddelen. Dat brengt een aantal technische problemen met zich mee, die in het beloofde volgende rapport uitgewerkt zullen worden. Een korte schets kan hier echter wel gegeven worden. Daartoe herinneren we aan wat eerder werd opgewerkt over de wijze waarop hier het cesuurprobleem geattaqueerd zou worden: het cesuurprobleem wordt in de eerste plaats beschouwd als een beslissingsprobleem, dat langs (zij het voorlopig ook heel eenvoudige) beslisskundige weg aangepakt kan worden (Raiffa 1968, Lindsey 1971). De cesuur moet zodanig gelegd worden dat onze beleidsdoelen op goede, zo mogelijk optimale wijze bereikt worden. Een belangrijk doel, misschien heeft dat zelfs de absolute prioriteit, is dat de studenten die een voldoende gaan krijgen ook op redelijke wijze tenminste aan de minimeisen voldoen. Een ander doel is dat zo weinig mogelijk studenten die aan die minimeisen voldoen op grond van feilbare toetsgegevens ten onrechte geen voldoende zouden krijgen. Een doel van weer geheel andere aard is dat de gehele procedure zo transparant mogelijk moet zijn voor de student, d.w.z. dat de student zo goed mogelijk in staat gesteld moet worden om zijn studiestrategie af te stellen op het behalen van een voldoende beoordeling. Een en ander houdt in dat het in het geheel niet een vanzelfsprekende zaak is om, wanneer minimeisen geformuleerd zijn, zeg in een simpel geval als: 80% van deze vragen moet je kunnen maken, deze eis zonder omwegen te vertalen als: op deze toets, die bestaat uit een willekeurige keuze van 50 vragen uit de lijst die we je aan het begin van de cursus hebben gegeven, moet je tenminste 40 vragen goed maken. Nee, voor het kiezen van de cesuur moet rekening gehouden worden met de toevalligheid van de keuze van de vragen uit de grotere lijst, met de invloed van allerlei toevallige, maar voor de beoordeling irrelevante, factoren op de studieprestatie van de student op dat moment; een correctie moet aangebracht worden op grond van het beleid dat de voorkeur geeft (zeg maar) aan het ten onrechte iemand een voldoende geven boven het ten onrechte iemand een onvoldoende geven; etcetera.

Het is niet denkbeeldig dat de docent die de hier voorgestelde cesuurbepaling in praktijk brengt, ontdekt dat binnen de in het studieprogramma daarvoor uitgetrokken tijd lang niet alle studenten, ook bij behoorlijke inspanning hunnerzijds, niet kunnen voldoen aan wat hij toch beschouwt als de allerminhaalste inhoudelijke eisen van stofbeheersing. Zou een dergelijke situatie aanleiding kunnen zijn tot het dan toch maar verhogen van de cesuur met als gevolg dat een belangrijk aantal studenten aan dit vak zeg twee keer zoveel tijd kwijt zullen zijn dan in het studieprogramma daarvoor uitgetrokken was? Of had er misschien een scherpere selectie vooraf moeten plaats vinden? Of moet de conclusie getrokken worden dat er kennelijk voor dit vak veel te weinig plaats is ingeruimd binnen het studieprogramma?

a) toch maar verhogen van de cesuur, en een herhaling instellen? Deze (on)mogelijkheid wordt hier volstrekt van de hand gewezen. Het is een onjuiste taakopvatting van de docent dat hij het niveau van zijn studieonderdeel op eigen houtje, buiten facultair overleg e.d. om zoals dat voorafgaat aan de vaststelling van studieprogramma en beleid, zou moeten handhaven door meer tijd van de studenten te vragen dan het studieprogramma daarvoor heeft vastgesteld.

b) moet er vooraf dan maar scherper geselecteerd worden? Dit is een mogelijkheid die overwogen kan worden. Een belangrijke opmerking is dat een dergelijke selectie aan hoge eisen van validiteit (en aanvaardbaarheid, De Groot 1972) moet voldoen. De eis van validiteit is de volstrekt redelijke eis dat een dergelijke selectie ook aantoonbaar moet werken zoals men haar bedoelt. In het onderhavige geval zou een scherpere selectie vooraf overwogen kunnen worden wanneer er een mogelijkheid is om met een dergelijke selectie ook inderdaad diegenen tegen te houden die ook bij behoorlijke inspanning hunnerzijds aan inhoudelijke minimeisen in de latere studie, minimeisen waarover binnen de subfaculteit overeenstemming bestaat, niet binnen redelijk tijdbestek kunnen voldoen. Anderzijds mogen niet tegengehouden worden degenen die zonder belangrijke problemen de vervolgstudie zouden kunnen doen.

Op meerdere plaatsen in dit rapport komt het belang van een zindelijke opvatting over het selectiegebeuren nog aan de orde. In dit geval kan aan selectie als oplossing van het probleem pas gedacht worden wanneer, zoals gezegd, tenminste binnen

de subfaculteit overeenstemming bestaat met betrekking tot de inhoudelijke minimeisen die men wil hanteren, en de maximale studietijd die men daarvoor ter beschikking wil stellen. Blijkt de afstemming van beschikbare tijd op de minimum prestatie zoals geëist niet goed, en selectie blijkt op behoorlijk wijze niet mogelijk, dan rest niets anders dan ofwel de eisen te laten zakken, ofwel meer tijd voor het betreffende vak uit te trekken.

c) moet er op subfaculteitsniveau onderhandeld worden over het tijdsbeslag voor dit vak en de inhoudelijk te stellen minimeisen?

Het antwoord hierop is duidelijk, gezien het onder b) besproken. Ja. Bovendien moet dit overleg noodzakelijk ook voorafgaan aan eventuele selectieve maatregelen. (Overigens, een selectieve maatregel die overwogen kan worden is bijvoorbeeld om de studierichting voor een qua vooropleiding minder diverse groep toegankelijk te stellen). Dergelijk overleg zou ook kunnen leiden tot geheel andere maatregelen, zoals opsplitsing van het betreffende vak in qua niveau verschillende 'stromen'.

Deze summiere bespreking maakt hopelijk duidelijk dat een hoop problematische onderwijssituaties in het huidige traditionele beoordelingsstelsel verduisterd kunnen worden. en in feite afgewenteld worden op de toch niet te brede rugen van studenten, waar het oneindig veel beter zou zijn om de problemen expliciet te onderkennen, en er een verantwoord onderwijsbeleid naar in te richten.

UITGANGSPUNTEN.

Kort besproken, en vergeleken met kerniternmethoden en methode Wijnen.

1e uitgangspunt is dat bij de oplossing van het cesuurprobleem expliciet van elkaar onderscheiden moeten worden:

het eigenlijke probleem van de cesuur tussen voldoende en onvoldoende de eventuele wenselijkheid om verschillende prestaties ook verschillend te 'belonen'; selectieproblemen; en problemen van 'externe' motivatie.

Dit is een uitgangspunt' dat voor de volledigheid maar even

uitgebreid behandeld wordt, maar eigenlijk een vanzelfsprekende zaak behoort te zijn. De Groot besteedde er in zijn presentatie van de kernitem. methode in 1964 de volgende woorden aan:

"Bij het samenstellen van proefwerken of examenopgaven en, vooral, bij het vaststellen van wat gewoonlijk de Inorment worden genoemd (welke prestatie leidt tot welk cijfer), zijn twee problemen aan de orde. die strikt gescheiden kunnen en moeten worden:

A. de differentiatie, de relatieve prestatiebeoordeling, d.w.z. de wijze waarop verschillen tussen en de rangorde van geleverde prestaties in schaal moeten worden gebracht;

B. de cesuur, de absolute prestatiebeoordeling, in termen van de dichotomie: voldoen of niet voldoen aan de eisen, voldoende of onvoldoende.

Deze twee aspecten van het beoordelingsprobleem hebben logisch niets met elkaar te maken. De vraag waar de cesuur moet liggen is principiëel onafhankelijk van de vraag in hoeverre men kan (of wil) differentiëren tussen prestaties. Ook als er een enorme differentiatie is. behoeft de grens tussen voldoende en onvoldoende het materiaal niet in tweeën te delen: in principe kan 'al het ingeleverde werk, hoe uiteenlopend van niveau ook, aan de gestelde eisen voldoen, dus 'voldoende' zijn; of ook: niet voldoen en dus zonder uitzondering 'onvoldoende' zijn."

De koppeling die traditioneel in ons onderwijs stelsel bestaat tussen cesuurbepaling en selectie is het onderwerp van herhaalde scherpe aanklachten van De Groot, zoals ook in zijn eerste deel van het tweeling artikel over de kernitemmethode (1964). In principe moet de onderwijsinstelling de verantwoordelijkheid voor een goede studievoortgang op zich nemen voor alle toegelaten studenten. Dat "op praktisch ieder niveau de mogelijkheid, dat een leerling geacht wordt te zijn 'afgefallen' als gevolg van 'ongeschiktheid' zodat de school zich dit afvallen niet als fout aanrekent" is een flagrante misstand.

Met name is het ook onjuist dat cesuurbepaling voor afzonderlijke vakken gekoppeld is aan examenregelingen, in de trant van: een

onvoldoende voor een bepaald vak betekent dat je gezakt bent voor het gehele examen (en allerlei misschien wat mildere varianten op dit grond thema). Dat is wat De Groot bedoelt met zijn stelling (1964):

"De in ons onderwijs veelvuldig voorkomende koppelings- en compensatieregelingen tussen vakken onderling maken een gezond cesuurbeleid per vak, goede, gefundeerde voldoende-onvoldoende beslissingen, vrijwel onmogelijk."

Gezien dit alles is het des te verwonderlijker dat De Groot met zijn uitwerking van de kernitemmethode aan dit grote bezwaar van het Nederlandse gebruikelijke beoordelingsstelsel in het geheel niet tegemoet komt. De kernitemmethode toepassen betekent op geen enkele wijze dat de docent zich hoeft te verzetten tegen het in zijn subfaculteit gebruikelijke selectie systeem (als je al van 'systeem' mag spreken).

Tweede uitgangspunt is dat de cesurbepaling plaats moet vinden op een wijze zodanig dat[^]de student zijn studie strategie er op af kan stemmen die voldoende ook in één keer te halen, ofwel: de toetsing op voldoende beheersing van de stof moet transparant zijn.

De student moet de beschikking hebben over een dermate precieze formulering van de minimum eisen, dat hij zich met die formulering als richtsnoer, en bij voorkeur geholpen door bijvoorbeeld tussentijdse toetsingen, zodanig kan voorbereiden op het afsluitende tentamen dat hij er zeker van kan zijn een voldoende daarvoor te zullen scoren. Bijkomende eis is dat die voorbereiding ook efficiënt kan zijn, d.w.z. dat de student zich op de minimum eisen kan voorbereiden zonder daarvoor nietnoodzakelijk onderwerpen te bestuderen, of onderwerpen in veel hogere mate te beheersen dan voor die voldoende prestatie noodzakelijk is.

De kernitemmethode van De Groot is in de praktijk in strijd met het uitgangspunt van transparantie. Docenten hebben onderling al de grootste problemen om uit te maken welke items tot 'kernitems' gepromoveerd moeten worden, laat staan dat de student een behoorlijk

inzicht heeft in de aard van de vragen aan de hand waarvan de cesuur vastgesteld gaat worden. Bij het 'berekenen' van de cesuur volgens de kernitemmethode zijn bovendien een aantal eigenschappen van de gebruikte kernitems van belang, waarover maar heel weinig en veel gevallen vrijwel niets bekend is vóórdat het tentamen is afgenomen. Dat de studenten van te voren op de hoogte gebracht worden dat voor de cesuurbepaling de kernitemmethode gebruikt zal worden, en welke variant daarvan precies door de docenten gekozen is, is een figuur waar ik nog nooit van gehoord heb; het zou bovendien weinig zinvol zijn omdat de student met een dergelijke mededeling alleen nog maar in groter onzekerheid komt te verkeren over waar hij nu precies met die eindtoets aan toe is.

De kernitemmethode in de variant van Van Naerssen is ongeveer het toppunt van gebrek aan transparantie. De docent wijst intuïtief in een rijtje items die op de toetsafname een hoge biseriële correlatie met totaalscore bleken te hebben, het item aan dat volgens hem het meest overeenkomt met wat hij denkt dat het minimum aanvaardbare niveau is. Uit wat de p-waarde van het aldus gekozen item op het al afgenomen tentamen bleek te zijn, wordt dan de aftestgrens afgeleid. De methode van Van Naerssen heeft als grote handicap dat ze voor vrijwel geen enkele docent begrijpelijk is, en zeker aan studenten niet begrijpelijk is te maken. Ook al zou het laatste lukken, dan is de student op geen enkele wijze in staat uit die informatie een studiestrategie af te leiden die hem rechttoe rechtaan naar een voldoende beoordeling kan voeren.

Van de methode Wijnen zou op het punt van transparantie gezegd kunnen worden dat de student die zich behoorlijk voorbereidt op een tentamen waarvan hij weet dat de methode Wijnen er bij gebruikt zal worden om de cesuur te bepalen, er redelijk zeker van kan zijn wel te zullen slagen. Het zou erg mooi zijn wanneer naast de mededeling dat de methode Wijnen gebruikt zal worden (en hoe die methode in elkaar zit en waar dat in de praktijk wat afwijzingspercentages betreft op neer komt) de student ook de beschikking krijgt over informatie met betrekking tot het soort vragen dat hij op dat tentamen mag verwachten, en hoe 'goed' dat soort vragen gemaakt pleegt te worden (bijv. afgaand op prestaties van vorige groepen studenten op het vergelijkbare tentamen). De methode Wijnen is een beetje lauw wat de

eis van transparantie betreft, maar voor een methode die bedoelt een ad-hoc oplossing voor het cesuurprobleem te bieden is dat oneindig veel beter dan het versluierende karakter van de pseudorationale kernitemmethoden.

overdoen afschaffen

derde uitgangspunt: zak-slaagbeslissingen moeten voldoen aan de normen van betrouwbaarheid en validiteit.

Dit is misschien het moeilijkste uitgangspunt, ook het moeilijkst toe te lichten. De achterliggende idee is eenvoudig genoeg: beslissingen over personen neem je niet zo maar, die moeten op deugdelijke gronden en procedures berusten. Beslissingen over personen moeten verantwoord kunnen worden tegenover de betrokken personen. Dat mag bij voorkeur zelfs geïnterpreteerd worden als volgt: dat beslissingen over personen in laatste instantie ook toetsing door de rechter moeten kunnen doorstaan. Het gaat hier onder andere inderdaad & m een stuk rechtszekerheid voor de student, Op die rechtszekerheid zal ik hier niet verder ingaan, omdat ik hier het accent niet wil leggen op de verantwoordbaarheid van de beslissingen naar de student toe, maar het kunnen verantwoorden van de genomen beslissingen tegenover het 'forum' van (in dit geval) psychologen onderwijskundigen. Vanwaar deze rare draai? Wel, ik zal wat ruimte besteden aan het uitleggen hiervan.

Het nemen van beslissingen over personen is in haar algemeenheid een nogal ongeregelde aangelegenheid, zij het dat voor bepaalde soorten van beslissingen men zich heeft te houden aan bepaalde rechtsregels. Bij het laatste is dan vooral het arbeidsrecht als voorbeeld te nemen. De meeste beslissingen over personen worden genomen door mensen die daarvoor geen enkele opleiding hebben ontvangen. die niet weten welke procedures zij het best kunnen volgen, welke informatie bruikbaar is en welke niet, en die bovenal voor hun werk geen normatieve en ethische richtsnoeren hebben meegekregen. Het nemen van beslissingen over personen is ook geen onderwerp dat zich in een grote belangstelling van wetenschapszijde heeft mogen verheugen. Met één grote uitzondering: bepaalde takken van

de psychologie hebben zich de laatste halve eeuw intensief bezig gehouden met de vraag hoe beslissingen beter genomen kunnen worden onder gebruikmaking van gegevens over verstandelijk en ander functioneren van de personen over wie die beslissingen te nemen zijn. Concreter: door gebruik te maken van gegevens die met behulp van psychologische tests verkregen kunnen worden, kunnen beslissingen over personen vaak beter genomen worden. Dat geldt voor bijvoorbeeld het sollicitatiebeleid van ondernemingen, maar ook voor het stellen van diagnoses in de medische zowel als de psychotherapeutische beroepspraktijk, als ook in het onderwijsbeleid waar het gaat om het vinden van de voor Jan of Marietje beste vormen van onderwijs. Hier, en in het volgende, moet de lezer zich terdege realiseren dat een studievoortgangstoets of een tentamen óók een vorm van psychologische test is, waarvoor dezelfde constructieve eisen gelden als voor tests in het algemeen gelden. Ik zal verder dan ook gewoon over 'toetsen' spreken. In de beginjaren van het toetstijdperk stonden voornamelijk constructieve en technische problemen in de aandacht, zij het dat ook al direct forse discussies losbrandden over de wenselijkheid van al die testerij. Naarmate het gebruik van toetsen en tests voor het nemen van beslissingen over personen op alle gebieden van het menselijk leven toenam, werd ook meer aandacht besteed aan normatieve eisen die aan de kwaliteit van de te gebruiken toetsen gesteld moeten worden. Over die normatieve eisen wordt nog steeds gediscussieerd, zij het dat langzamerhand de meningsverschillen gereduceerd zijn tot meer subtiele kwesties, en er wat de dagelijkse toetspraktijk betreft overeenstemming bestaat ten aanzien van te hanteren normen. Die discussie over, en formulering van, normen voor test en toets gebruik heeft voornamelijk plaats gevonden binnen de grote Amerikaanse associaties voor psychologen en onderwijskundigen. Deze associaties hebben de normen voor toetsing beschreven in 'Standards for Educational and Psychological Tests and Manuals,' laatste herziene editie 1974, uitgegeven door de American Psychological Association. Het zijn deze 'Standards' die ik zal gebruiken om cesuurbepalingsprocedures op hun aanvaardbaarheid te toetsen.

Het derde uitgangspunt kan dan ook als volgt geformuleerd

worden: zak-slaagbeslissingsprocedures moeten voldoen aan de Standards.

De plotselinge overgang van normen voor toetsen naar normen voor beslissingsprocedures vraagt om toelichting. Traditioneel dacht men nogal eens dat zaken als betrouwbaarheid en validiteit en zo (termen die ik hier niet zal toelichten, maar die later nog uitgebreider aan de orde kunnen komen) kwalitatieve eigenschappen van een toets konden zijn, dus los van de situatie waarin die toets gebruikt werd, en los van het doel waarvoor die toets gebruikt werd. Het denken over betrouwbaarheid en validiteit van toetsen en tests evolueerde erg snel, en in de Standards valt nu dan ook te lezen dat de kwaliteit van een toets afhankelijk is van de situatie waarin, en het doel waarvoor ze gebruikt wordt. Of met andere woorden: er is niet zoiets als de validiteit of geldigheid van een bepaalde toets, maar van beslissingen die op basis van de toets resultaten genomen worden. Vandaar dat ik in het volgende consequent zal proberen te schrijven over betrouwbaarheid en validiteit van beslissingsprocedures, en niet van de daarvoor gebruikte toetsen* En omgekeerd; beslissingsprocedures waarin toetsresultaten een belangrijke rol spelen, zullen moeten voldoen aan de Standards hoewel deze Standards nog steeds in belangrijke mate beschreven zijn als te gelden voor toetsen.'. In de laatste editie van de Standards is een hoofdstuk toegevoegd dat dit veranderd taalgebruik weerspiegelt: dat handelt namelijk over de toepassing van toetsen, en niet meer over toetsen 'sec' (hoewel ook in het voorafgaande deel van de Standards voortdurend er op gewezen wordt dat kwaliteitsonderzoek van een toets alleen maar kan plaats vinden in dezelfde of strikt vergelijkbare situaties als waarin de toets gebruikt zal gaan worden).

Een toetsing van bekende cesuurmethoden tegen de Standards kan op deze plek niet anders dan heel grof gebeuren. Wat betrouwbaarheid betreft, kan van de kernitemmethoden zonder meer gezegd worden dat ze onbetrouwbaar zijn, dat wil zeggen dat de gehele beslissingsprocedure van begin (formulering van items waaruit kernitems aangewezen worden) tot het eind (berekening van de cesuur op grond van de verkregen toetsingsgegevens) bijzonder vatbaar is voor allerlei onbedoelde en toevallige invloeden. Ofwel:

mensen rond de aftestgrens kunnen er van verzekerd zijn dat ze geslaagd of gezakt zijn dankzij een samenloop van omstandigheden. Het verschijnsel hangt ook samen met het volledige gebrek aan transparantie van de toetsingssituatie bij deze kernitemmethoden. Het zou erg verhelderend kunnen werken een analyse van het begrip betrouwbaarheid te maken in termen van transparantie van de toetsingssituatie. Met betrouwbaarheid zou je bijvoorbeeld kunnen bedoelen dat de student tijdig voldoende zekerheid moet hebben over hoe de zak-slaagbeslissing voor hem uit zal vallen. Een beslissingsprocedure zou dan betrouwbaar genoemd kunnen worden naar de mate waarin de student terecht die zekerheid over zijn al dan niet zullen slagen, kan hebben.

De methode Wijnen houdt rekening met een bepaald soort onbetrouwbaarheid van de toets (de 'homogeniteit' van de toets), die van invloed is op de meetkwaliteit van de toets, zoals uitgedrukt in de standaardmeetfout. Deze standaardmeetfout is immers de correctiefactor die Wijnen (ten gunste van de student) toepast (tweemaal die standaardmeetfout) op zijn 'referentiepunt': de gemiddelde prestatie. Als geheel een beetje een twijfelachtige procedure, waar betrokkenen echter op de korte termijn vrede mee zouden kunnen hebben.

De validiteitsproblematiek ligt een stuk complexer dan de betrouwbaarheidsproblematiek. Wat de methode Wijnen betreft, moet helaas opgemerkt worden dat Wijnen zich met validiteitsvragen in het geheel niet heeft bezig gehouden, behalve door op te merken dat validiteit van toetsing een probleem voor de docent is, en daar moet een buitenstaander (psycholoog of onderwijskundige) zich niet mee bemoeien! Commentaar overbodig.

De kernitemmethoden zitten wat validiteit betreft ook bijzonder moeilijk. Allereerst kan opgemerkt worden dat volledig in de kernitemmethoden ontbreekt dat de cesuurbeslissingen een bekende voorspellende geldigheid hebben, bijvoorbeeld ten aanzien van het met succes verder kunnen studeren. O.K., het is misschien ook helemaal niet de bedoeling geweest van De Groot en Van Naerssen om een procedure te ontwikkelen die studenten die in de verdere studie

moeilijkheden zouden ondervinden door hun onvoldoende beheersing van dit bepaalde vak, tegen te houden. Maar dan kan het niet anders of de kernitemmethoden moeten aan eisen van 'inhoudelijke' validiteit voldoen. Ook daar ontbreekt echter bij deze methoden iedere garantie voor, omdat uiteindelijk de docent de kernitems aanwijst (volgens De Groot) of de minimumgrens (volgens de gewijzigde methode Van Naerssen) op basis van zijn niet verder geëxpliciteerde opvattingen daarover (intuïtie kun je rustig zeggen). Op grond van welke overwegingen de docent bepaalde Items als kernitems kan aanwijzen, en andere niet, daarover zwijgt De Groot in alle talen. (dat een kernitem de kern van de stof moet betreffen, en andere even vage algemeenheden, biedt de docent geen enkele steun).

vierde uitgangspunt.

Cesurbepaling is niet zo maar een activiteit die los staat van de context van het onderwijs. Ieder voorstel voor een methode om cesuren te bepalen of te kiezen zal in de onderwijs situatie aanvaardbaar moeten zijn. Bij voorkeur wordt met de cesurbepaling een bijdrage geleverd aan het bereiken van de doelstellingen die met het onderwijs nagestreefd worden; tenminste zal een cesurbepalingsaanpak niet met belangrijke doelstellingen van onderwijs strijdig mogen zijn.

Een bijna vanzelfsprekende en daarom vrijwel nooit expliciet genoemde doelstelling van onderwijs is het geven van onderwijs. Verre van een grapje, deze uitspraak, want het ziet er in de dagelijkse onderwijspraktijk nogal eens naar uit dat niet het geven van onderwijs, maar het sorteren van studenten de belangrijkste bezigheid van onderwijsinstellingen is. Hoewel selectieve maatregelen op bepaalde momenten beat noodzakelijk of gewenst kunnen zijn, heeft het geven van onderwijs zélf niets met selectie te maken. Of er al dan niet vooraf of achteraf geselecteerd wordt, zodra er onderwijs gegeven wordt, kan het niet anders of het moet de bedoeling zijn met dat onderwijs ook alle tot dat onderwijs toegelaten studenten te onderwijzen.

Voorgaande gemeenplaatsen leiden tot de interessante conclusie dat bij een behoorlijke opzet van het onderwijsprogramma, en een

behoorlijke inzet van de studenten, aan het eind van de rit deze studenten alle een voldoende moeten verdienen.

Belangrijkste uitgangspunt bij het zoeken naar een methode voor het bepalen van cesuren is dan ook dat in principe iedere student die het aangeboden onderwijs behoorlijk heeft gevolgd, een voldoende krijgt. In de praktijk zijn er altijd een klein aantal uitzonderingen: degenen die met reden niet aan het onderwijs konden deelnemen, een enkele pechvogel die door misverstanden de boot mist, een enkeling die ten onrechte tot dit stuk onderwijs toegelaten werd.

Het is erg belangrijk zich voortdurend te realiseren dat cesuurbepaling los gezien moet worden van eventuele selectieve maatregelen, evenals het niet meer dan zinnelijk is om het geven van onderwijs te zien als een activiteit die op zich niets met selectie te maken heeft. Voor het hier geformuleerde uitgangspunt dat in principe iedereen voor een afsluitende toets hoort te slagen betekent dit, dat het ook onder het in praktijk brengen van dit principe mogelijk blijft om bepaalde studenten of groepen studenten met minimaal voldoende resultaten voor een bepaald vak, of voor alle vakken, door middel van een gekozen selectiebeleid uit de vervolgstudie te weren.

Het is heel waarschijnlijk dat de eerste reactie op de formulering van dit principe er één is in de trant van: maar dat is gruwelijk onmogelijk voor het onderwijs dat ik geef! Er is alle kans dat hier niet het probleem ligt bij het principe zoals boven geformuleerd, maar bijvoorbeeld het voor een bepaald vak veel te krap bemeten zijn van de in het studieprogramma beschikbaar gestelde tijd, of het politieke feit dat tot een bepaalde studie nu eenmaal studenten moeten worden toegelaten met zeer wijd uiteenlopende vooropleidingen.

De Groot zou het liefst meegaan met het principe dat een onvoldoende niet aan studenten gegeven hoort te worden, maar aan de kwaliteit van het gegeven onderwijs; hij merkt daarbij echter op dat het Nederlandse gebruik van cijfergeven een dergelijke opvatting in de weg staat, en legt zich bij deze stand van zaken neer: "Onvoldoende" zou moeten betekenen: Bij deze leerling heeft mijn onderwijs gefaald

maar deze (absolute) betekenis is haast niet te handhaven als de officiële cijferschaal suggereert dat een onvoldoende net zovn 'gewoon' cijfer is als een voldoende en dat onvoldoendes in liefst vijf verschillende gradaties kunnen voorkomen. (Paedagogische Studiën 1964).

De kernitemmethode zoals gepresenteerd door De Groot in 1964 gaat er van uit dat die studenten een voldoende verdienen die de kern van de stof begrepen hebben. De kern van de stof is die verzameling van kennis, inzichten en vaardigheden die van essentieel belang zijn voor het kunnen begrijpen van de stof (het theoretisch inzicht), bij latere toepassing van het geleerde, e.d. Kortom, hoewel de omschrijving veel te vaag is, valt wel ongeveer te begrijpen wat De Groot hier bedoelt.

De Groot ontlokt de docent een formulering van minimum eisen in absolute termen, waar het duidelijk is dat de minimum eisen die je aan de student kunt stellen, zullen variëren al naar gelang de voor dit vak beschikbare tijd, de omvang van de gepresenteerde stof en de moeilijkheid daarvan, en de kwaliteit van de studenten (kwaliteit dan opgevat naar de variatie in capaciteiten en in voorkennis).

Op zich is er natuurlijk alles v66r om bij de cesuurbepaling, uit te gaan v66n de verzameling minimale doelstellingen die bereikt moeten zijn, en dat is dan ook de aanpak die ik verder zal gebruiken, zolang daarbij maar voorop staat dat dergelijke minimum eisen volledig afhankelijk zijn van wat in een bepaalde onderwijs situatie haalbaar is.

Blijkt dat niet alle studenten aan bepaalde geformuleerde eisen kunnen voldoen, dan betekent dat de noodzaak om de minimum eisen anders te formuleren, het onderwijs te verbeteren, meer tijd voor dit vak in het onderwijsprogramma uit te trekken, of andere denkbare maatregelen te treffen zolang het probleem maar niet opgelost wordt door gemakshalve dan maar wat meer studenten een 'onvoldoende' te geven.

De gewijzigde kernitemmethode van Van Naerssen gaat zonder verdere omwegen er van uit dat de docent in een verzameling items die items kan aanwijzen dit 'zijns inziens het moeilijkheidsniveau

markeren van deze ondergrens (d.i. de cesuur tussen voldoende en onvoldoende, b.w.) van beheersing van de stof'. De kernitemmethode zo opgevat is eigenlijk al lang geen methode voor cesuurbepaling meer, omdat op het kritieke moment de docent op grond van zijn vage ideeën maar een gok moet doen, terwijl juist van een cesuurbepalingsmethode verwacht mag worden dat zij de docent in staat stelt zijn ideeën over waar de cesuur zou moeten liggen duidelijk gestalte te geven.

Ook Wijnen besteedt, evenals De Groot, veel aandacht aan de uitgangspunten op grond waarvan mede een cesuurbepalingsmethode vorm gegeven moet worden. Hij noemt in zijn 'Explicitering van een onderwijsopvatting' vier punten, waarvan er drie in overeenstemming zijn met het door mij hier geformuleerde uitgangspunt dat in principe iedere student in één keer hoort te kunnen slagen, terwijl het vierde uitgangspunt een meer technische kwestie behelst (de norm dat bij beslissingen over personen rekening gehouden moet worden met de kwaliteit van de informatie waarop die beslissing gefundeerd is, als ik het in iets andere bewoordingen mag stellen dan Wijnen doet in zijn punt 3 op blz. 51):

1. Uitgangspunt is de stelling, dat onderwijs afgestemd behoort te zijn op de mogelijkheden van degenen, die tot het volgen van dat onderwijs gekwalificeerd zijn. Dit impliceert, dat de te behandelen stof wat betreft moeilijkheid, tempo van behandeling en omvang rekening moet houden met de te onderwijzen groep#

2. De in het onderwijs bereikte resultaten worden opgevat als resultanten van verschillende factoren. Tenminste moeten worden onderscheiden: aan de vragen gebonden factoren en aan de personen gebonden factoren. Welke factor in hoofdzaak een eventueel gunstig of ongunstig resultaat heeft veroorzaakt, kan zonder nadere bewijsvoering nauwelijks worden vastgesteld. Het is in ieder geval onaanvaardbaar, indien men een neiging, die gunstige én ongunstige resultaten zonder meer aan de leerlingen toeschrijft, kritiekloos zou volgen.

4. Omdat onderwijs in de huidige organisatievorm gegeven wordt

aan groepen van leerlingen en omdat onderwijs afgestemd dient te zijn op die groepen, komt een afwijzing of uitsluiting eerst dan in aanmerking, wanneer met een voldoende mate van waarschijnlijkheid kan worden vastgesteld, dat iemand ten onrechte deel uitmaakt van de groep, waarvoor het onderwijs bestemd is."

Kritische vraag is echter of Wijnen er in geslaagd is zijn methode een vorm te geven die in overeenstemming is met dit uitgangspunt. Daar kan verschillend over gedacht worden: een welwillende interpretatie, die ook in de lijn ligt van het denken van Wijnen zelf over het cesuurprobleem, is dat we op korte termijn zo weinig van de relevante onderwijsprocessen te weten kunnen komen dat we in tussentijd maar een aanvaardbare gok moeten doen en het gemiddelde van de prestaties als referentiepunt nemen, en daar twee maal de standaardmeetfout van de studietoets van af moeten trekken om een redelijke cesuur te krijgen. Op de wat langere termijn bezien, is de methode Wijnen een aanpak die heel weinig te maken heeft met de door hem geformuleerde uitgangspunten, omdat tenslotte de keuze van het gemiddelde als referentiepunt niet dwingend beargumenteerd kan worden, in feite zelfs principieel onjuist is omdat altijd het gewenste minimum niveau van beheersing van de stof als referentiepunt gekozen moet worden. Het bezwaar tegen de laatste aanpak is voornamelijk van praktische aard: voorlopig zijn de technische problemen die daarbij komen kijken, nog niet opgelost óf wordt een onderwijsopzet en uitwerking vereist die behoorlijk wat tijd vraagt voordat ze rond is. vandaar ook dat de methode Wijnen als adhoc cesuurbepalingsmethode grote verdienste heeft.

Vijfde uitgangspunt: Kiezen uit alternatieve procedures om de doelstellingen van cesuurbepaling zo goed mogelijk te realiseren.

Bij cesuurbepaling gebruik je niet zo maar de eerste de beste methode die iemand daarvoor suggereert. Een behoorlijke methode voor cesuurbepaling is afhankelijk van de doelen die met die cesuurbepaling gediend moeten worden. Geenszins een open deur, deze uitspraak. Vraag eens een collega docent wat eigenlijk zijn bedoelingen zijn bij het vaststellen van de zakslaag grens voor tentamens, en je hebt grote kans hem met de mond vol tanden te

betrappen.

Het zou misschien ook een veel betere aanpak zijn om die hele cesuurbepalingsproblematiek maar eens de deur uit te gooien, en opnieuw beginnen te denken vanuit de doelstellingen van het onderwijs, zowel wat de globale, maatschappelijke doelstellingen betreft als de inhoudelijke, de situatie van dat onderwijs, het soort studenten dat dat onderwijs volgt, de diverse materiële en immateriële randvoorwaarden waarbinnen het onderwijs gerealiseerd moet worden, etcetera. Dan stuit je vanzelf een keer op zoiets als een cesuurprobleem voor een bepaald vak, omdat ergens een grens getrokken moet worden tussen wat binnen dit onderwijs stelsel nog als een voldoende prestatie mag worden beschouwd, en welke prestatie dat niet meer is.

Ook daarmee zijn we er nog niet, omdat voor het trekken van die grens tussen voldoende en onvoldoende in principe een oneindig aantal verschillende mogelijkheden bestaan, en dan vooral een groot aantal alternatieven die een veranderde onderwijs opzet behelzen. Bijvoorbeeld: wanneer toetsing op dit moment 'geplaagd' wordt door het verschijnsel dat grote aantallen studenten kennelijk voor het tentamen opkomen alleen om het een keer te proberen, dan ligt het voor de hand het cesuurprobleem hier niet op te lossen door gebruikmaking van kernitemmethoden, methode Wijnen, of besliskundige benaderingen (zie volgende paragraaf), maar moet een oplossing kennelijk gezocht worden in de richting van het meer transparant maken van de toetsings situatie, meer spreiden van toetsing over het hele onderwijs in plaats van een grote hoeveelheid stof in een klap te willen vragen, e.d.

Kortom, eerst en vooral gaat het er om een inventarisatie te maken van de vele mogelijkheden die er zijn zowel op het punt van de inrichting van het onderwijs, als waar het meer gaat om toetsingsmethoden zelf, mogelijkheden die instrumenteel zijn voor het bereiken van die doelstellingen die de docent (de faculteit) bij de cesuurbepaling voor ogen staan. Uit deze verschillende alternatieven kan een keuze gedaan worden, o.a. door af te checken tegen eerder genoemde uitgangspunten, door ze te vergelijken op het punt van

de mate waarin belangrijke onderwijsdoelstellingen er mee bevorderd worden, en op meer prozaïsche aspecten als de mate van praktische realiseerbaarheid (qua kosten en tijdsbeslag bijv.).

Ik hoef niet veel woorden te besteden om duidelijk te maken dat noch de kernitemmethoden, noch de methode Wijnen, m.b.t. dit uitgangspunt kunnen voldoen. De kernitemmethode-auteurs gaan er zonder omwegen vanuit dat deze methode altijd van toepassing (de 'beste' aanpak?) zijn, in welke specifieke onderwijs situatie men ook zit, en wat ook de specifieke doelstellingen van dat onderwijs zijn. De methode Wijnen gaat er van uit dat juist over die onderwijssituatie weinig bekend is, maar dat betekent nog geen vrijbrief om dan maar niet meer te zoeken naar alternatieven die in ieder geval op een enkele doelstelling van dat onderwijs pertinent beter zijn dan het hanteren van de methode Wijnen.

Uitgangspunt zes.: het aantal foutieve beslissingen wordt zo klein mogelijk gehouden, met name ook rekening houdend met verschillen in 'ernst' van verschillende soorten foutieve beslissingen over individuen.

Stel eens, dat in min of meer ideale formulering scherp omschreven is welke mate van beheersing van de stof als onvoldoende aangemerkt moet worden, en welke als voldoende. De werkelijkheid ziet er altijd minder mooi uit dan de verbale wereld, en zeker wanneer we op het punt aanlanden dat het niveau van beheersing gemeten moet gaan worden weten we vrij zeker dat die meting zo onnauwkeurig zal zijn dat een relatief belangrijk aantal studenten ten onrechte een voldoende of onvoldoende beoordeling dreigt te krijgen. Ik zie er nu even van af dat het uitgangspunt dat toetsing en de daarop te volgen beslissingsprocedure transparant moet zijn, ook ten doel heeft om de student te beschermen tegen de risico's ten onrechte als onvoldoende beoordeeld te worden.

Iedere methode voor cesuurbepaling levert ergens een aftestgrens op, en het is altijd zeker dat onder hantering van die aftestgrens een aantal studenten ten onrechte een voldoende beoordeling zullen krijgen, en bij niet volstekte (een onbereikbaar ideaal?) transparantie

dat ook een aantal studenten ten onrechte een onvoldoende zullen krijgen. Hoewel in een bepaalde situatie natuurlijk niet bekend is welke studenten ten onrechte een voldoende hebben gekregen (als we dat wisten zaten we niet net dit probleem), moet het wél zo zijn (Standards!) dat bekend is om welke aantallen foute beslissingen van iedere soort het ongeveer gaat. En ook moet bekend zijn, hoe die aantallen foutieve beslissingen veranderen, wanneer de aftestgrens iets verschoven wordt, ofwel omlaag, ofwel omhoog. Wel, als dat het geval is, dan is het duidelijk dat hier een mogelijkheid aanwezig is om de cesuur *optimaal, te kiezen, d.w.z. net zo lang te verschuiven totdat we schatten een aantal fouten te maken dat minimaal is.

Op zich is deze werkwijze weinig opzienbarend, zij het dan dat de methode zonder uitzondering bij de cesuurbepaling in het hoger onderwijs tot nu toe niet wordt toegepast. De zaak wordt interessanter op het moment dat een foute beslissing niet meer louter als 'fout' beschouwd en geteld wordt, maar er onderscheid wordt gemaakt in categorieën van foute beslissingen. In het onderwijs zal men bijvoorbeeld in het algemeen er veel meer voor beducht (moeten) zijn studenten ten onrechte een onvoldoende te geven, dan studenten ten onrechte een voldoende te geven. Deze notie kan wat preciezer geformuleerd worden, en het verschil in 'gewicht' van beide soorten foute beslissingen kan gekwantificeerd worden en meegenomen bij de procedure om een 'optimale' aftestgrens te zoeken. Niet alleen dat het allemaal mogelijk is is hier aan de orde, het gaat er om dat het zo ook gedaan moet worden, en dat een behoorlijke cesuurbepalingsprocedure ook uitdrukkelijk op dit uitgangspunt gebaseerd is. Daarbij blijft natuurlijk een moeilijkheid dat dit uitgangspunt maar tot op bepaalde hoogte te realiseren zal zijn zonder in strijd te komen met andere uitgangspunten, en met het nog te formuleren uitgangspunt van de kostenbeheersing. Maar dat is een afzonderlijk probleem, waarvoor ik later zal proberen mogelijke oplossingen aan te geven.

Ook ten aanzien van dit uitgangspunt moet weer geconstateerd worden dat de kernitemmethoden er niet op gebaseerd zijn, en er op geen enkele wijze rekening mee houden. De docent die de kernitemmethode (of dat nu de oorspronkelijke of de variant van Van

Naerssen is) hanteert, voert in feite een magisch ritueel uit (althans voorzover het dit uitgangspunt betreft): hij weet niet wat het resultaat er van is. Niet alleen weet de docent niet of zijn uiteindelijke aftestgrens bij benadering 'optimaal' is wat het aantal foutieve beslissingen betreft, hij weet zelfs niet of zijn aftestgrens wat dit betreft ook maar de naam 'aanvaardbaar' verdient.

Wat de methode Wijnen betreft gaat het er op lijken dat ik in herhalingen verval, maar toch moet ik ook hier weer zeggen dat deze ad hoc procedure van cesuurbepaling althans in zoverre rekening houdt met dit uitgangspunt dat de methode er op gericht is om het aantal 'ernstige beslissingsfouten', d.w.z. ten onrechte uitgedeelde onvoldoendes, klein te houden. Hoewel ook bij de methode Wijnen onbekend blijft of de uiteindelijk er uit rollende cesuur bij benadering optimaal is wat de som van foutieve beslissingen betreft, wordt er wel geprobeerd om te waarborgen dat er in ieder geval geen grote afwijkingen in ongewenste richting voorkomen. Jammer genoeg zijn er in de onderwijspraktijk nogal eens docenten die wel met Wijnen mee willen gaan in de keuze van het gemiddelde als referentiepunt, maar slechts één keer de standaard meetfout als correctie daarop willen hanteren: zij lopen een groot risico een onaanvaardbaar hoog aantal onvoldoendes uit te delen aan studenten die hun zaakjes in feite wel behoorlijk beheersten.

Bij het tegen elkaar afwegen van de ernst van een foutief gegeven onvoldoende tegenover een ten onrechte gegeven voldoende, loopt de docent een erg groot risico daarbij al te gemakkelijk af te gaan op een aantal populaire, vooropgezette, opvatting over hoe het onderwijs functioneert, en welke studenten daarin goed uit de bus zullen komen en welke niet. Er kan nooit genoeg de nadruk op gelegd worden, en ik zal dat met name ook in het rapport met de presentatie van de methodieken doen. dat allerlei opvattingen pas van invloed mogen zijn op het te kiezen beleid wanneer voldoende empirische of theoretische ondersteuning voor die opvattingen aanwezig is. Een voorbeeld is de mate waarin iemand die een bepaald vak zogenaamd 'onvoldoende' zou beheersen, daardoor en daardoor alleen in de rest van de studie gehandicapt zou zijn wanneer hij het niet over zou hebben gedaan. Dit soort uitspraken zijn niet van de categorie waar de docent het zich bij

kan veroorloven af te zien van kritisch onderzoek van de houdbaarheid van de uitspraak in deze bepaalde onderwijs situatie. De moeilijkheid met dit soort opvattingen is echter dat de docent niet op de hoogte is, en misschien is het ook teveel om van hem te verwachten daarvan op de hoogte te zijn, van de resultaten van onderzoek en onderwijskundige theorievorming voorzover ze licht kunnen werpen op de houdbaarheid van zijn opvattingen. Ik zal dan ook proberen om de meest voor de hand liggende missers te anticiperen, en te behandelen,

Voor het overige kan ik de docent alleen maar aanraden om zijn eigen opvattingen over onderwijs en over capaciteiten van studenten even kritisch te beschouwen als uitspraken over samenhangen op zijn eigen vakgebied: "Wat bedoel ik hier precies mee?", "Op welke waarnemingen is die opvatting gebaseerd?" "Is dat ooit in mijn onderwijs situatie onderzocht en getoetst?" "En als het nu eens inderdaad juist zou blijken te zijn, wat heeft het dan nog voor beleidsconsequenties?" (wat het laatste betreft: ook opvattingen over het soort maatregelen dat je zou moeten nemen wanneer bepaalde verschijnselen zich voordoen, moeten heel kritisch bekeken worden; niet in het minst omdat misschien erg voor de hand liggende maatregelen vaak een tegengestelde uitwerking blijken te hebben dan er van verwacht werd).

Resumerend. De uitgangspunten waaraan cesuurbepalings methoden bij voorkeur moeten voldoen, of in overeenstemming waarmee ze geconstrueerd moeten worden, zijn:

1. Het cesuur probleem wordt gescheiden gehouden van ' andere problemen als selectie, differentiële beoordeling, motivatie.
2. Toetsing op voldoende beheersing van de stof moet transparant zijn; ook de beslissingsprocedure moet transparant zijn.
3. De beslissingsprocedure moet voldoen aan normen zoals gesteld in de Standards. (betrouwbaarheid en validiteit van de beslissingen).
4. In principe krijgt iedere student die het onderwijs behoorlijk volgt, een voldoende.
5. De keuze van te volgen beslissingsprocedure vindt plaats op

grond van behoorlijke afweging van alternatieve mogelijkheden.

6. Is de procedure eenmaal vastgesteld, dan wordt de grens zakken slagen voor die procedure op besliskundige wijze bepaald (geoptimaliseerd).

Een methode voor het bepalen van de cesuur is aanvaardbaar in de mate waarin aan genoemde uitgangspunten voldaan is. Tenminste moet van een cesuurbepalingsmethode duidelijk zijn óf, en zo ja in hoeverre, zij aan ieder van de uitgangspunten voldoet; omdat deze informatie op zich nog weinig zinvol is, moet ook bekend zijn waarom een bepaalde methode slechts in bepaalde mate aan een bepaald uitgangspunt voldoet, welke pogingen ondernomen zijn om de methode te verbeteren, op welke punten de methode eventueel verbeterd kan worden, zij het mogelijk tegen verslechtering van het kunnen voldoen aan andere uitgangspunten (welke tradeoffs tussen verschillende uitgangspunten mogelijk zijn), en zo mogelijk ook hoe deze methode zich ten aanzien van de uitgangspunten laat vergelijken met andere bekende of tot nu toe gebruikte methoden.

Uitgangspunten, besproken en geconcretiseerd.

Afzonderlijke aanpak voor selectie en andere problemen.

Een verantwoorde aanpak van het cesuurprobleem vraagt dat het ontdaan wordt van 'vreemde' elementen als selectieve bedoelingen e.d. Een belangrijke vraag met betrekking tot de haalbaarheid van mijn cesuurbepalings-alternatief is dan ook in hoeverre die scheiding te maken is, en in de praktijk vruchten afwerpt. Maar niet alleen voor de cesuurbepaling is het belangrijk om het onderscheid tussen verschillende soorten problemen te maken. Ook wat betreft het bereiken van selectie doelen is het noodzakelijk om de selectie los te koppelen van cesuurbepaling voor afzonderlijke vakken. Ten aanzien van het uitgangspunt loskoppeling van cesuurbepaling en selectie, zal bij bespreking van de overige uitgangspunten gedemonstreerd worden hoe die cesuurbepaling aangepakt kan worden zonder daarbij tegelijk selectief bezig te zijn. Voor de aanvaardbaarheid van de door mij voorgestelde cesuurbepalingsaanpak is het niet zo gek wanneer ik op

deze plaatseen korteschets geef van de wijze waarop het selectie gedoe aangepakt kan worden los van de cesuurbepaling die gebezigd wordt.

Allereerst kunnen we samen toch wel vaststellen dat er in iedere studie richting binnen het hoger onderwijs geselecteerd wordt. Niemand kan staande houden dat alle studiestakers tot hun beslissing gekomen zijn omdat ze kennelijk te lui waren om behoorlijke studieresultaten te bereiken. (Met de categorie serieuze omzwaaiers wil ik hier geen rekening houden, omdat ik aanneem dat de redenen voor omzwaaien van studierichting gelijk zullen zijn bij hantering van andere cesuurbepalingsmethoden; het oorzakelijk verband tussen studieprestatieniveau en veranderen van studie acht ik met andere woorden klein genoeg om het hier te mogen verwaarlozen).

Voor het in elkaar zetten van selectie procedures zijn de volgende punten van overwegend belang:

1. wat wil men met die selectie bereiken;
2. hoe denken verschillende partijen over 1);
3. wie bepaalt in feite de wijze waarop selectie plaats vindt (d.i. welke belangen groep heeft de grootste zeggenschap);
4. welke belangen van welke andere betrokkenen zijn 'onbeschermd';
5. wordt selectie gerechtvaardigd op grond van een te klein (absoluut) aantal beschikbare plaatsen (voor studenten in de opleiding, of afgestudeerden in de samenleving); of
6. wordt selectie beargumenteerd op gronden van 'niveau bewaking';
7. op grond van welke informatie wordt selectie van studenten bepaald;
8. wat is de rechtvaardiging voor het gebruik van juist die informatie, en voor de wijze waarop van die informatie gebruik gemaakt wordt;
9. een beschrijving van de volledige 'beslissingssituatie' of de volledige beslissingsstrategie;
10. in hoeverre is aantoonbaar dat de beslissingsstrategie in overeenstemming is met het doel van de selectie zoals omschreven in 1).

Een puntsgewijze toelichting:

1. Wat wil men met die selectie bereiken.

In het algemeen zijn doelen van selectie in het onderwijs zeer onuitgesproken, dat wil zeggen niet expliciet geformuleerd. Desgevraagd is een docent bijv. altijd wel bereid om te verklaren dat de selectie doeleinden nobel en nastrevenswaardig zijn. Het is natuurlijk niet voldoende alleen maar het idee te hebben een juist selectiebeleid te voeren: ten minste zal een expliciete formulering van dat beleid, en de doelen die er mee bereikt moeten worden, nodig zijn. Bij pogingen tot formulering van beleid en na te streven doeleinden kan al snel blijken dat een en ander toch niet zo vanzelfsprekend is als men eerst oprecht meende. Maar een dergelijke 'ontdekking' is bijzaak; waar het om gaat is dat selectie procedures alleen zinvol geconstrueerd kunnen worden wanneer bekend is welke resultaten er van verwacht worden. Controle op de aanvaardbaarheid van selectieprocedures is ook slechts mogelijk wanneer dat kan gebeuren door de resultaten van de gehanteerde procedures te vergelijken met de bedoelde resultaten. (Daarnaast zal altijd bekeken moeten worden of de selectie aanvaardbaar is vanuit ethische overwegingen, en vanuit normen voor verantwoord gebruik van informatie zoals vastgelegd in de Standards (1974).

De doeleinden die men nastreeft hangen af van de belangen die men bij die selectie heeft: in punt 2 komt aan de orde hoe die belangen kunnen verschillen. Of de selectie voorafgaand aan het onderwijs plaats vindt, of 'gedurende de rit', soms zelfs pas achteraf. Of het gaat om het zo goed mogelijk vullen van slechts een beperkt aantal beschikbare plaatsen, of om het handhaven van een bepaald 'niveau'; of men met zo weinig mogelijk onderwijs een klein aantal studenten tot hoog niveau wil opvijzelen, of met veel ondersteunend onderwijs juist een grote categorie studenten tot bepaalde minimum niveaus van vakbeheersing wil brengen (selectie is vaak gericht op het weren van studenten van wie op een of andere manier 'minder' verwacht wordt, maar natuurlijk zou selectie ook in enigszins omgekeerde zin gebruikt kunnen worden als het uit het dagonderwijs weren van studenten

van wie men mag aannemen dat zij zonder intensieve begeleiding ook uitstekende prestaties kunnen behalen). Deze opmerking tussen haakjes is, wederom tussen haakjes, ook in ander opzicht interessant: je zou kunnen vermoeden dat bij het chaotische selectieproces zoals dat nu nog in het hoger onderwijs voorkomt, het resultaat is dat juist onderwijsvragende studenten afgeschrikt worden, en studenten die het ook zónder begeleiding prima aankunnen aangemoedigd worden door te studeren: je zou je kunnen afvragen of het een behoorlijke doelstelling voor een onderwijsinstelling kan zijn de selectie te richten op het zo moeiteloos mogelijk een aantal niet onderwijsvragende studenten naar het doctoraal te 'sluizen'.

Wanneer doelstellingen voor selectie geformuleerd zijn, en stel nu eens dat men expliciet gekozen heeft voor het selecteren tijdens de rit van het onderwijs (tegen uitdrukkelijke en herhaalde protesten van onderwijskundigen in), dan kan vervolgens de vraag gesteld worden in hoeverre de huidige laissezfaire selectie gewoonten voorzien in die doelstellingen, en in hoeverre alternatieve procedures daar misschien beter in zouden kunnen voorzien. En het zal een bijzonder harde dobber worden om te demonstreren dat het huidige chaotische selectie stelsel waar niemand enige greep op heeft, op de bedoelde wijze, en geen andere, uitwerkt.

2. verschillende doelen voor verschillen belangengroepen.

Er is een jammerlijke tendens in het denken over selectie om er daarbij zonder meer van uit te gaan dat uitsluitend de belangen van de selecterende instantie relevant zijn. Een uitgebreide illustratie daarvan is te vinden in de discussie over toelatingsregelingen voor numerus fixusstudierichtingen: de al te grote vanzelfsprekendheid waarmee woordvoerders uit of leden van betrokken studierichtingen selectie voorstellen verdedigen die (uitsluitend) in het belang van de betrokken faculteit zijn.

Op zijn minst gaat het bij iedere selectie echter om twee partijen: de selecterende en de geselecteerde partij. Daarnaast kan het in veel gevallen ook nog voorkomen dat er andere belangen eveneens op het

spel staan. Een illustratie van het laatste: wanneer bij numerus fixus regelingen grote aantallen studenten gedwongen worden om in plaats van hun studie van 1e keuze, een studie van 2e keuze te doen, is het evident van groot belang voor die studierichtingen die in belangrijke mate als 2e voorkeur gekozen worden, op welke wijze de numerus fixus uitgevoerd wordt: als een zeer goede selectie op algemene capaciteiten (zoals aangegeven door eindexamencijfer gemiddelde bijv.) mogelijk zou zijn en uitgevoerd wordt, betekent het dat faculteiten die in grote mate door numerus fixus afgewezenen gekozen worden, geconfronteerd worden met een aanbod van studenten met kwalitatief geringe studieverwachtingen. In een dergelijk geval (wat zich op het ogenblik niet voordoet omdat juist op basis van die eindexamencijfers zo beroerd weinig te voorspellen valt t.a.v. toekomstig studiesucces) botsen de belangen van numerus fixus studierichtingen met die van nog lopen' studierichtingen.

Daarnaast is het duidelijk dat de verantwoordelijkheid voor de kwaliteit en inrichting van het onderwijs ligt bij het kabinet, gecontroleerd

door de volksvertegenwoordiging: het is geenszins vanzelfsprekend dat de belangen van een bepaalde faculteit dezelfde zijn als het 'algemeen belang'.

Doelen van selectie, en selectieprocedures, zullen dus verschillen al naar gelang zeg de studenten, de faculteit, of overkoepelende beleidsinstanties daar de dikste vinger in hebben.

In de praktijk zal gezocht moeten worden naar een doelen pakket, en een inrichting van de selectie procedure, die voor alle betrokken partijen 'aanvaardbaar' is (zie De Groot 1972).

3. Wie bepaalt in feite de selectie gang van zaken.

In de praktijk loopt het allemaal niet zo mooi, en heeft de faculteit waarschijnlijk de grootste invloed op de selectie. Tenzij die selectie zo volstrekt ongecontroleerd verloopt als nu gebruikelijk is, dan heeft niemand daar enige greep op, en mag je rustig aannemen dat alle betrokken partijen door dat uitblijvende beleid

in hun belangen geschaad worden. Maar ook al verloopt de huidige 'methode' van selecteren volstrekt chaotisch, toch blijft de verantwoordelijkheid voor die selectie bij de faculteit: zij is aansprakelijk te stellen voor alle onzorgvuldige beslissingen (in dit geval meer: het uitblijven van verantwoorde beslissingen) over individuele studenten. En dan hebben we het maar liever helemaal niet over de veel ingewikkelder maatschappelijke belangen die door verregaande onachtzaamheid bij beoordeling en selectie in het hoger onderwijs geschaad worden (de onnodige studieduurverlenging als gevolg van de beoordelingsgewoonten die we van onze voorouders hebben overgenomen zijn een punt in kwestie).

4. Onbeschermd belangen van de student.

Omdat selectie in het huidige hoger onderwijs op de slordigst denkbare manier plaats vindt, zodat je zelfs niet kunt spreken van beslissingen die genomen worden, verkeert de student die deze vloed van ellende over zich heen krijgt, in een beklagenswaardige en tenminste onbeschermd positie. Tegen beslissingen die niet genomen worden kun je nergens in beroep gaan, gesteld al dat we in het huidige stelsel over behoorlijke beroeps procedures zouden beschikken. Het meest haalbare is af en toe een voorzichtig protest tegen een individuele vakgroep die het bij haar zakslaag beleid kennelijk al te bont heeft gemaakt (afwijs percentages buiten de meest extreme grenzen van ieder denkbare vorm van redelijkheid), of handelt in tegenstelling tot eerder gedane toezeggingen (dan wil een administratiefrechtelijke procedure nog wel eens een mogelijkheid voor de benadeelde student zijn).

Ook wanneer het selectiebeleid wél behoorlijk in elkaar zit, en expliciet is geformuleerd, zal zich nog al eens de feitelijke situatie voordoen dat het beleid sterk eenzijdig door de faculteit bepaald wordt. Dan is het zaak dat de faculteit, of andere overkoepelende beleids instanties, de belangen van studenten behartigen: dat er behoorlijke procedures voor beroep op bepaalde beslissingen bestaan, dat er controle mogelijk is op de 'kwaliteit' van de selectieprocedure (bijv. de mate waarin voldaan is aan ethische en technische normen,

waarbij wat de laatste betreft de Standards als leidraad zouden kunnen dienen). Een bescherming van de rechtspositie van de student analoog aan het arbeidsrecht (door De Groot bij herhaling bepleit) is bijna helemaal te mooi om waar te zijn.

5. Quota selectie.

Een belangrijk onderscheid tussen verschillende vormen van selectie is dat tussen selectie voor een vastgesteld en te klein aantal plaatsen, en selectie op 'kwaliteit'. De laatste vorm van selectie, waarbij in principe het aantal beschikbare plaatsen niet van invloed is op de selectie procedure (totdat bijv. economische overwegingen nopen die ideale toestand te wijzigen), komt in de volgende paragraaf aan de orde.

Zijn er voor een bepaald aantal kandidaten slechts een beperkt aantal plaatsen beschikbaar, dan is er sprake van quota selectie. Een bekend voorbeeld daarvan is selectie voor numerus fixus studierichtingen. Quota selectie is de vervelende opgave om uit een groep in principe geschikte of aanvaardbare kandidaten een selectie te maken voor het te kleine aantal beschikbare plaatsen. In de praktijk bestaat de moeilijk onderdrukke neiging om in dergelijke gevallen uit het 'aanbod' de 'besten' te selecteren. Dergelijke praktijken zijn niet vanzelfsprekend ook maatschappelijk 'gewenst', of ethisch aanvaardbaar, althans kunnen daarover de meningen verdeeld zijn. Uiterst bedenkelijk wordt het wanneer quota selectie, indien al noodzakelijk in het onderwijs, wordt uitgevoerd tijdens en niet voorafgaand aan het onderwijs: dat komt er op neer dat bijzonder hooghartig en al te makkelijk beschikt wordt over jaren van studie-inspanning door afgewezen studenten geleverd. Ik neem dan ook maar liever aan dat geen zinnig mens een dergelijke wijze van selecteren voorstaat, en dat iedere vorm van selectie tijdens het onderwijs, op zich al treurig genoeg dat we dat nog steeds tolereren, een selectie op kwaliteiten is.

6. Selectie op bewezen of verwachte kwaliteiten.,

De vorm van selectie zoals die voor onderwijs meest karakteristiek is, is selectie op bekwaamheden, ongehinderd door overwegingen van

beperkte plaatsingsmogelijkheid. Althans, het is expliciet de bedoeling van onderwijs dat een ieder het best bij hem passende onderwijs kan volgen.

Het is makkelijk om een dergelijke ideale scheiding tussen beide soorten selectie te maken, maar in de praktijk zal het vaak zo zijn dat zogenaamde kwaliteitseisen of bekwaamheidseisen zich al snel aanpassen aan dát niveau waarmee de beschikbare plaatsen over het algemeen het prettigst gevuld worden. En omgekeerd: selectie louter op kwaliteiten zal er op den duur ook in resulteren dat een bepaald aantal plaatsen beschikbaar komt (wanneer bijvoorbeeld door de jaren heen de fluctuatie in het op deze wijze toegelaten aantal studenten niet erg groot is), zodat vervolgens het risico bestaat dat selectie zodanig plaats gaat vinden dat het beschikbare aantal plaatsen ongeveer gevuld wordt: kwaliteitsselectie degenereert dan (misschien geheel onbedoeld, maar toch) tot quotaselectie. Het laatste zal sneller gebeuren wanneer kwaliteits- of bekwaamheidseisen niet duidelijk te stellen zijn, wat bij hoger onderwijs typisch het geval is omdat niemand behoorlijk inzicht heeft in welke eisen het beroep aan de academicus stelt, zo er al sprake is van een bepaald beroep (ook voor exacte vakken is het aantal beroepsmogelijkheden na de studie zeer uiteenlopend van karakter en inhoud), en al aangenomen mag worden dat die eisen door de jaren heen min of meer van dezelfde aard zullen blijven.

Het is duidelijk dat in een situatie als deze, bij selectie tijdens of na afloop van het hoger onderwijs het risico bijzonder groot is dat de selectie er louter één is in een soort pikorde van knapheid. Ik bedoel daarmee te zeggen dat niemand in staat is om ook maar de geringste overtuigende argumentatie voor dit soort selectie aan te voeren (zonder terug te vallen op overwegingen die de selectie herleiden tot een vorm van quota selectie, in welk geval alle moeite gedaan moet worden om die selectie nu juist voorafgaand aan het onderwijs uit te voeren). Ik heb natuurlijk geen enkele moeite met di vormen van studievoortgangs bewaking die moeten voorkomen dat bepaalde lieden al lanterfantend of louter op hun dikke zitvlees de papieren in de wacht slepen omdat er een bepaalde hoeveelheid tijd verstreken is sinds 'aanvang' van de studie, Maar ook bij de laatste vorm van selectie moet aantoonbaar zijn, aannemelijk gemaakt kunnen worden, dat de

selectie inderdaad functioneert zoals ze bedoeld is te functioneren.

Over diverse vormen van selectie is erg veel bekend wat betreft technieken en hulpmiddelen die daarbij te gebruiken zijn, normen die aan het gebruik van tests, toetsen en andere informatie daarbij gesteld moeten worden (Standards 1974), etcetera etcetera. De lezer die zich op dat gebied wil informeren kan zich makkelijk de relevante literatuur verschaffen (vooral het boek van Thorndike (ed.) *Educational Measurement* is een bijzonder goede introductie).

7. Op welke gegevens over de student wordt er geselecteerd?

Ook al vindt selectie volstrekt chaotisch en ongecontroleerd plaats, dan nog kan bekeken worden op welke gronden studenten voortijdig de studie worden gedwongen te verlaten. Voor het krampachtige Nederlands in de eerste zin bestaat in het Amerikaans taalgebruik een fraai eufemistische term: 'cooling out'. De constatering dat studieprestaties achterblijven bij die van andere studenten is behoorlijk machteloos: natuurlijk zijn er altijd studenten die het om wat voor redenen dan ook minder goed doen dan anderen. Kritische vraag bij het uitselecteren van studenten is niet of ze het minder goed doen dan een bepaald percentage anderen, maar of ze niet voldoen aan bepaalde behoorlijk geformuleerde minimum eisen.

Wie een voorbeeld kent van een studierichting waarin studenten de studie staken omdat ze aan dergelijke minimum eisen niet kunnen voldoen, en niet omdat ze niet zo goed zijn als een aantal andere studenten, wordt vriendelijk maar dringend verzocht zich met mij in verbinding te stellen: ik heb zo iets nog niet ergens kunnen constateren. Het zou al bijzonder mooi zijn wanneer selectie op een dergelijke wijze expliciet gemaakt kon worden; lukt dat, dan kun je meteen opmerken dat deze vorm van selectie tijdens het onderwijs bijzonder ongelukkig is, omdat je toch bij voorkeur zou moeten proberen studenten die niet aan de minimum eisen kunnen voldoen al te signaleren voordat ze aan de studie toegelaten worden.

Op het ogenblik vindt selectie plaats op basis van behaalde

studieprestaties waarvan alleen maar gezegd kan worden dat ze slechter zijn dan die van een aantal andere studenten. De student die met de studie ophoudt ziet waarschijnlijk weinig mogelijkheden om in redelijke tijd aan de exameneisen te voldoen. Wat studenten een 'redelijke' tijd vinden, zal heel erg uiteen lopen, je mag misschien wel veronderstellen dat juist de 'goede' studenten door studievertraging sterker afgeschrikt worden om door te gaan dan studenten die de wereld wat minder regel bekijken. Met andere woorden: het systeem van op grote schaal laten overdoen van studieonderdelen zou op zich een selectieve werking kunnen hebben die omgekeerd is aan wat men bedoelt. Maar goed, dit is niet meer dan een veronderstelling. Wat veel duidelijker is, is dat het 'systeem' alleen belang hecht aan de cijfertjes, en niet aan de tijd die aan het behalen van die cijfers besteed is. Voor een selectie die bedoeld is een selectie op 'kwaliteiten' te zijn, is dat een volstrekt absurde zaak. Dezelfde studieprestaties voor verschillende studenten zijn met elkaar volstrekt onvergelijkbaar wanneer studenten zeer uiteenlopende hoeveelheden studietijd daarin hebben gestoken, en men daarmee bij examens (bij de selectie dus) geen rekening wil houden. Het ligt voor de hand een dergelijke procedure te rechtvaardigen met het soort argumenten als: je moet studenten toch in de gelegenheid stellen bepaalde missers te herstellen. Dat is juist, alleen komt er natuurlijk wel ergens een grens waar beneden je misschien nog van 'missers' mag spreken, maar waar over heen je toch moet spreken van een systematisch tekort schieten van de capaciteiten van de student. Zéker in die gevallen waarin selectie gerechtvaardigd wordt met een verwijzing naar het geven van een garantie voor een behoorlijke beroepsuitoefening, zou het handhaven van het huidige systeem op zijn best halfslachtig genoemd kunnen worden. Wil men daarentegen uitsluitend de selectie laten bepalen door de hoeveelheid kennis en inzicht die men tijdens de studie opdoet, zonder beroep op validiteit daarvan naar de latere beroeps situatie toe, dan is er juist weer geen reden om een beleid te voeren waarin studenten in grote mate stof zitten te herkauwen die ze al eens eerder hebben moeten bestuderen.

Een behoorlijke selectieprocedure kan dan ook niet anders als rekening houden met de bestede studietijd. Het best zou zelfs zijn om te proberen die studietijd zo gelijk mogelijk te houden voor alle

studenten: wanneer selectie uiteindelijk moet gebeuren door toetsing van prestaties van de student aan bepaalde minimum normen? heeft dat alleen zin wanneer de voorbereidingstijd van de student binnen redelijke grenzen beheerst wordt. Een voorbeeld kan dat misschien beter verduidelijken. Wanneer voor de toelating tot numerus fixus studierichtingen het behalen van hogere eindexamencijfers de kansen verhoogt, dreigt de misstand te ontstaan dat leerlingen in het v.w.o. met opzet het laatste jaar gaan doubleren om tot die hoge eindcijfers te kunnen komen. Het spreekt vanzelf dat naarmate een dergelijk verschijnsel grote omvang zou aannemen, ieder positief verband dat voordien misschien nog bestaan zou hebben tussen gemiddeld eindexamen cijfer en latere studieprestaties, geringer zal worden. Weer terug vertaald: wil een bepaalde selectieprocedure op studieprestaties voldoen aan redelijke validiteitseisen, dan moet een minimale controle op de studietijd benodigd voor het behalen van die studieprestaties, aanwezig zijn.

Het is dan alleen maar prettig wanneer langs geheel andere wegen ook nog aangetoond kan worden dat beheersing van de studietijd van de student efficiënter en onderwijskundig meer verantwoord is dan het vrij laten daarvan zoals in het huidige systeem, en ook weer onder de herstructurering het geval is.

Wanneer selectie eindelijk weer langs fatsoenlijke weg gebeurt, en selectie beslissingen dan ook expliciet door de faculteit genomen worden, dan ontstaan ook mogelijkheden voor het bij die beslissing betrekken van andere soorten informatie dan alleen studieprestaties. Allereerst zijn natuurlijk bepaalde persoonlijke omstandigheden van belang (ziekte e.d.). Daarnaast komt dan allereerst in beeld het testen van capaciteiten. Of hier nog praktische mogelijkheden liggen voor het tertiaire onderwijs is sterk de vraag, gezien het al sterk geselecteerde karakter van deze groep studenten; maar in principe zou je kunnen overwegen om niet alleen de behaalde studieprestaties, maar ook de capaciteiten ('begaafdheden') van de student bij de beslissing te betrekken. We hebben hier immers ongeveer hetzelfde als met die studietijd: hoe langer iemand nodig heeft gehad om een bepaald prestatieniveau te bereiken, des te minder waarde kan daaraan gehecht worden. Voor de student die minimale prestaties levert, en voor wie

selectie overwogen wordt, is het kennelijk van belang te weten of hij prestaties levert die beneden zijn vermogen liggen, of dat hij met zijn geleverde prestaties al op de toppen van zijn kunnen zit. Wanneer selectie gebeurt op basis van verwachtingen over studiesucces in de volgende studiefase, is het van evident belang voor de selecterende instelling om dat te weten. Voor de student geldt dat hij zich onterecht behandeld kan voelen wanneer hij op grond van een marginale prestatie uitgeselecteerd wordt, maar van zichzelf weet dat hij een behoorlijk stuk beter kan presteren (je zou kunnen zeggen dat deze student dan een ongelukkige studiestrategie heeft gevolgd: maar dat is een iets ander verhaal, dat thuishoort onder het hoofdstukje transparantie van beslissingsprocedures.)

Je zou nog een paar stappen verder kunnen gaan, en ook biografische gegevens, en persoonlijkheids- en motivatie gegevens bij de selectie procedure betrekken: daarmee zouden we ons dan wel op ethisch bijzonder glad ijs begeven. Maar toch, je zou wel degelijk moeten onderzoeken in hoeverre dergelijke informatie over studenten in verband staat met selectieve beslissingen die over hen genomen worden, en wel voornamelijk om de vinger te kunnen leggen op onbedoeld en ongewenst functioneren van de selectie. Evaluatie van selectie procedures zal dan ook zeker dit soort onderzoek moeten omvatten.

Het spreekt vanzelf dat de selectie zoals die al dan niet bedoeld plaats vindt, in overeenstemming moet zijn met de algemene doelstellingen van onderwijs. Dat met name ook de informatie over de student op basis waarvan geselecteerd wordt, in bepaalde relatie moet staan tot die doelstellingen. Dat ook de selectie procedure op zich in overeenstemming moet zijn met die doelstellingen. Op diverse plaatsen in dit rapport heb ik al voldoende aandacht aan dit soort minimum voorwaarden besteed, om er hier niet nog verder over uit te hoeven weiden.

Een voorstel.

Tot nu toe heb ik nog heel weinig gezegd over de mogelijkheden om een meer behoorlijk selectie stelsel in te richten. Ik zal daar heel kort

over zijn. In de Verenigde Staten kent men in het hoger onderwijs een cijferstelsel dat zich in principe zou kunnen lenen voor een behoorlijke opzet van het hele beoordelingssysteem. Natuurlijk, aan dat Grade Point Average stelsel kleven ook wel enige bezwaren, maar het is in de verste verte niet zo slecht als het fossiele en versteende stelsel zoals we dat in Nederland kennen. Voor een beschrijving van het GPA systeem verwijs ik naar Bevez, die ook enkele onderzoek resultaten bespreekt. Beoordeling van studieprestaties vindt meestal plaats op een vijfpuntsschaal, waar alleen de laatste plaats gereserveerd is voor wat een 'onvoldoende prestatie' genoemd mag worden: onvoldoendes plegen dan ook in aanzienlijk mindere mate te vallen dan wij gewend zijn. De cijfers voor een bepaald vak worden 'gewogen' naar gelang de tijdbesteding voor dat vak (studietijd die er in het studieprogramma voor is uitgetrokken) eventueel ook de zwaarte of moeilijkheid. De gewogen cijfers voor de verschillende vakken worden gemiddeld, en het zo verkregen Grade Point Average is één van de gegevens op grond waarvan de student tot vervolg studie (graduate school bijv., of het tweede jaar van college) toegelaten wordt. De studenten zijn op de hoogte van de normen m.b.t.

GPA die bij selectieve beslissingen gehanteerd worden, en kunnen zich in hun studie dan ook naar vermogen daar op baseren (je kunt a.h.w. een goed Grade Point Average 'opbouwen'). 0 ja, de te besteden studietijd wordt strak in de hand gehouden, geen gerotzooi dus met overdoen van belangrijke aantallen vakken.

Kenmerk van het Grade Point Average stelsel, waartegen voorzover ik weet geen belangrijke bezwaren ooit ingebracht zijn, is dat je in vergaande mate voor slechte prestaties op een bepaald vak kunt compenseren door betere prestaties op eventueel meer andere vakken. De Nederlandse docent is wat al te makkelijk geneigd om ineens geweldig principieel te gaan doen zodra er voorstellen aan de orde zijn om bijvoorbeeld in het kader van een examenregeling een grotere mate van compensatie van onvoldoende cijfers mogelijk te maken (let wel, 'onvoldoende' in het gebruikelijke beoordelingsstelsel, wat slechts zoveel wil zeggen als: slechter dan een x aantal andere. Ik vermoed dat ik over het compenseren van cijfers nog wel een afzonderlijk

hoofdstukje in elkaar mag zetten,

omdat mijn cesuurbepalingsvoorstel ook lijnrecht leidt naar een stelsel waarin op ruime schaal compensatie gelegenheid geboden wordt (zij het dat het dan geen compensatie van 'onvoldoende' prestaties hoeft te zijn, maar compensatie van 'minimale' prestaties. En inderdaad zou je kunnen zeggen dat mijn alternatieve voorstel slechts neer komt op het verwisselen van woorden met de bedoeling om van de slechte en absurde connotaties en van een stuk archaïsch woordgebruik af te komen).

Hoe valt transparantie te realiseren?

Er zijn twee verschillende zaken die ieder voor zich transparant moeten zijn, en die we dan ook afzonderlijk moeten behandelen. Ik bespreek allereerst transparantie van de toets zelf, en kom pas na behandeling van de andere uitgangspunten en hun realisatie of concretisering tot een cesuurbepalingsmethode, terug op die andere transparantie, namelijk van de beslissingsregel zelf. De wijze waarop de cesuur bepaald zal worden uitgaande van de toetsscores, ofwel de manier waarop voor de afzonderlijke student op grond van zijn scores bepaald wordt of zijn prestatie voldoende is of niet, moet transparant zijn, wat wil zeggen dat het de student niet alleen duidelijk moet zijn hoe de cesuur bepaald wordt en hoe bepaald wordt of zijn prestatie daar onder of daar boven ligt, maar bovendien moet de beslissing zo in elkaar zitten dat de student zich bijtijds kan voorbereiden op het behalen van een prestatie die als tenminste voldoende uit de bus zal komen. Maar de manier waarop dit in zijn werk kan gaan, wordt zoals gezegd pas veel later besproken.

Het is geen eenvoudige zaak om de toetsing transparant te maken voor de student, En dat is het zeker niet wanneer de docent meent voor die toetsing gebruik te moeten maken van het meerkeuze vraag 'format'. Ik kom daar direct nog op terug, want ik maak de opmerking eigenlijk om de lezer er vast op voor te bereiden dat hij straks een korte uiteenzetting krijgt waarom in de meeste onderwijs situaties de vorm van de meerkeuze toets beter vermeden kan worden.

Transparantie van de toets komt er op neer dat de student weet welke soort vragen hij voorgeschoteld zal krijgen. Transparantie betekent vanzelfsprekend ook dat tentamenvragen geen verrassing meer mogen zijn, zoals in het geval dat er een soort vraag gesteld wordt die in het onderwijs niet aan de orde is geweest, of dat er vragen gesteld worden waar ineens een stuk informatie of slimheid voor nodig is waar ook in het onderwijs niet naar toe gewerkt is. Ik hoop dat iedereen hier nog zonder enige bedenking met mij mee kan gaan. Het wordt al wat meer gevraagd, wanneer ik de lezer er op wijs dat het nog al op grote schaal voorkomt dat in een toets vragen gesteld worden waarvoor de student stukken kennis met elkaar moet combineren op een wijze zoals ook in het onderwijs niet aan de orde is geweest. Om een voorbeeld te nemen: stel het tentamen gaat over persoonlijkheidstheorieën, zeg van Freud tot Cattell. Persoonlijkheidstheorieën gaan over van alles en nog wat, en dat is voor verschillende persoonlijkheidstheorieën altijd wel ongeveer hetzelfde zij het dat de kleine verschillen nu juist de bestaansreden van nog weer een nieuwe persoonlijkheidstheorie plegen te zijn. Het is een gruwelijk onmogelijke opgave voor een student om in een afzienbaar tijdbestek op ieder mogelijk onderdeel iedere twee persoonlijkheidstheorieën met elkaar te vergelijken: het aantal mogelijke combinaties is gewoon te groot. Met als gevolg dat iedere tentamenvraag die wél een dergelijk soort vergelijking van de student vraagt, daarom eerder thuis hoort in een intelligentie test of voor mijn part een t.v. quiz, dan op een behoorlijke eindtoets. Maar hoe zit het dan, hoe kom je aan behoorlijke vragen voor een eindtoets, hoe formuleer je die?

Doorzoeken van de literatuur levert het ongelooflijk teleurstellende resultaat op dat er wel hele ritsen aanwijzingen gegeven worden welk soort vraagformuleringsfouten je zoal kunt maken, en welke trucs er zijn om een stuk tekst om te zetten tot een aantal toetsvragen, maar er is vrijwel nergens iets te vinden over hoe je op een theoretisch en onderwijskundig gezonde manier vragen moet construeren die in overeenstemming zijn met de doelstellingen van onderwijs. Stel dan bovendien nog als eis dat alleen die vraag constructie methoden bruikbaar zijn die ook voor de student inzichtelijk zijn, en door de student zélf toegepast zouden kunnen worden, en je zit volledig met

de handen in het haar omdat er dan uit de literatuur helemaal geen aanwijzingen meer komen.

Om, die onverwachte leemte op te vullen heb ik een poging gedaan een aantal vraag constructie regels op te stellen die theoretisch verantwoordbaar kunnen zijn, onderwijskundig zinvol zijn, voor de student inzichtelijk zijn, en door de docent toegepast kunnen worden bij het snel en op grote schaal formuleren van zinvolle vragen. Deze vraag constructie regels zijn niet voor ieder soort onderwijs noodzakelijk geschikt, maar wél naar het zich laat aanzien voor die soort leerstof die zich zinvol laat toetsen door tentamens samengesteld uit een redelijk aantal afzonderlijke vragen. Daarom hier een korte schets van dergelijke regels.

De veronderstelling die ik maak, is dat de leerstof zich op laat delen in de afzonderlijk behandelde begrippen, en de behandelde relaties tussen begrippen. Relaties tussen begrippen noem ik regels. Wat begrippen zijn? Ik noem er een paar: tafel is een begrip, stoel, bed, blinde darm ontsteking, kapitaal, steekproefverdeling, validiteit, studietoets, etcetera. Wat regels zijn? Dat zijn (wetmatige) samenhangen tussen begrippen: $E=mc^2$, de wet van de verminderende meeropbrengst; spreiding van tentamenscores; historische samenhangen; oorzaak en gevolg verbanden; etcetera. In veel gevallen is het onderscheid tussen een begrip en een regel niet erg duidelijk, of hangt het van het niveau waarop de stof behandeld wordt af: ik hecht niet aan een eenduidig onderscheid tussen beide, het onderscheid is in de praktijk alleen maar nodig om enige mogelijke verwarring te vermijden.

De manier waarop je een begrip 'beheerst', 'je eigen gemaakt hebt' is voor velerlei variatie vatbaar. Er zijn situaties waarin het voldoende is dat de leerling als een papegaai de definitie van een begrip moet kunnen afdraaien (zijn die situaties er nog?); belangrijker is dat de student voorbeelden van een begrip kan thuis brengen, een tafel als tafel herkent. Het herkennen van een tafel als tafel is vaak niet erg eenvoudig: op de kleuterschool zal het nog lastig genoeg zijn om gewone tafels in alle verscheidenheid als tafels te laten herkennen. Maar er zijn ook heel wat voorwerpen die niet evident zich als tafel of

niettafel laten classificeren. Ik laat het verder maar aan kunstenaars en anderen over om dat soort onduidelijke voorwerpen te ontwerpen; het wordt pas echt menens wanneer bijvoorbeeld voor het vak histologie de student voor hem bijzonder subtiele verschillen in structuur en kleur van preparaten moet kunnen gebruiken om juiste labels aan die preparaten te kunnen hangen. Zo zijn er op ieder vakgebied moeilijke classificatieproblemen die op zich onderwerp van onderwijs zijn. Nog een paar voorbeelden: opgaven als het evalueren van integralen laten zich in verschillende soorten opdelen al naar gelang het soort oplossingsmethode dat bruikbaar is; experimenten zijn er in soorten al naar gelang het soort data analyse techniek dat er op van toepassing is.

Voor veel soorten leerstof is het mogelijk de stof op te sommen in een lijst van begrippen en regels die daarin aan de orde komen. Bij ieder item uit die lijst zou aangegeven moeten worden op welk niveau dat begrip of die regel beheerst moet worden, d.w.z. welk soort vragen er over beantwoord moet kunnen worden. (Voor een gedetailleerde lijst van mogelijkheden zie mijn voorlopige verslag van februari/maart 1976). Op dit punt aangeland zie ik de bedenkingen van de docent levensgroot worden: dat is me nogal niet wat, dat vraagt een hoop schrijfwerk bijna net zo uitgebreid als mijn hele syllabus! Ik zou hier dan ook de docent willen herinneren aan het op de vorige bladzijde besprokene, dat nogal eens van de student op het tentamen gevraagd wordt iets geheel nieuws met de stof te doen, een nieuw soort kunstje te vertonen. Met name in die situaties waar de docent meent dat volledig uitschrijven van zijn leerstof onderdelen en de wijze waarop de student ze moet beheersen, een onbegonnen werk is, juist daar zal het ook voor de student een 'onbegonnen werk' zijn zich behoorlijk op het tentamen voor te bereiden zolang hij niet over een dergelijke lijst de beschikking heeft.

Een enkele uitzondering valt natuurlijk wel te maken, zoals voor die vakken waar van de student gevraagd wordt een zeer omvangrijke hoeveelheid literatuur door te lezen. Maar ook dan blijft het zaak de toetsing afgestemd te houden op de bedoelingen van het onderwijs, en op de eis van transparantie e.d., zoals onder andere zou kunnen gebeuren door in dergelijke gevallen gebruik te maken van een soort

toetsvragen die door Hofstee 'onbenullige items' werden gedoopt, en die niet meer pretenderen te doen dan te toetsen of de student de stof tenminste een keer doorgenomen heeft.

Het maken van een lijst begrippen en regels met bijbehorende gevraagde niveaus van beheersing kan natuurlijk eenvoudig zijn in die gevallen waarin de begrippen en regels van ongeveer gelijke 'orde' zijn, of zich in zulke groepen laten samenbrengen. Dan nog zal het nodig zijn om bij ieder item uit de lijst voorbeelden van vragen te geven, die de beheersing op het niveau zoals aangegeven toetsen,

Voor het formuleren van de minimum eisen voor een vak zijn er verschillende mogelijkheden. Die minimum eisen kunnen geformuleerd worden als een bepaald deel van de stof dat tenminste beheerst moet worden op het niveau zoals in de lijst aangegeven. Andere mogelijkheden zijn er natuurlijk ook, bijvoorbeeld door voor ieder item uit de lijst, ieder begrip of regel, verschillende niveaus van beheersing te specificeren: een niveau van beheersing waarop met dit onderwijs 'gemikt' wordt, een niveau van beheersing waaraan de student tenminste moet voldoen voor de kwalificatie 'voldoende beheersing van de stof', e.d. Een en ander hangt af van de aard van het onderwijs, de stof die behandeld wordt, toetsingsmogelijkheden, etcetera.

Het is dus mogelijk om de toetsing transparant te maken. De docent zal er wat moeite voor moeten doen, maar daar staat tegenover dat er nog enige extra payoff van zijn inspanningen is: verheldering van zijn doelstellingen en vooral van de leerstof zoals hij die in syllabi, op college en in werkgroepen presenteert kan ook voor hem (en natuurlijk voor de student) erg groot zijn; misschien het prettigst voor de docent is dat hij nu op snelle wijze een groot aantal items kan maken voor de toetsen die hij wil gebruiken, en dat hij die items (voor of) na de toetsafname ook gewoon vrij kan geven omdat het maken van nieuwe items een kleinigheid zal zijn. De studenten kunnen dan beschikken over een grote 'pool' van items van het soort als in hun eigen eindtoetsing ook aan de orde zullen komen.

De wijze waarop de stof in de toets gerepresenteerd zal zijn is

natuurlijk ook belangrijk vanuit transparantie overwegingen. Het is erg voor de hand liggend dat de student volledig op de hoogte is van welke gedeelten van de stof niet in de eindtoetsing aan de orde zullen komen, welke onderdelen een belangrijke plaats op de eindtoets zullen krijgen ofwel daarin zelfs zeker aan de orde zullen komen, en welke iets minder belangrijke onderdelen met elkaar strijden om in een klein aantal vragen aan de orde te komen.

Meerkeuze vragen of niet? Het probleem met meerkeuze vragen is, wat transparantie betreft, dat ze vrijwel niet transparant te krijgen zijn, althans dat daar veel meer moeite voor nodig is dan bij de huis tuin en keuken openeind vragen. Door de keuze van afleiders kan een meerkeuze vraag zeer veel moeilijker gemaakt worden dan een open eind vraag, en de docent weet vaak in de verste verte niet hoe het zit met de moeilijkheid van een bepaalde meerkeuze vraag, terwijl hij dat wel zou weten bij een open eind vraag. Afgezien van de op zijn best twijfelachtige transparantie die met meerkeuze vragen te bereiken is, kleven er aan meerkeuze vragen een groot aantal andere nadelen, die niet door voordelen geneutraliseerd worden, althans niet in de meeste onderwijssituaties waar de docent zijn eigen studietoets maakt voor een niet zo reusachtig aantal studenten.

Cesurbepaling_opzetten in overeenstemming met de Standards.

De Standards (1974) zijn niet zomaar een collectie regels en voorschriften aan de hand waarvan je zou kunnen nagaan of bepaalde procedures wel aan behoorlijke technische normen voldoen. De Standards vormen vooral een collectie relevante, en soms wat minder relevante, overwegingen van belang bij het opzetten van beoordelingsprocedures. Ook hier geldt dat we niet passief achteraf bepaalde voorstellen voor cesurbepaling toetsen tegen de normen zoals vastgelegd in de Standards, maar dat geprobeerd zal worden om een cesurbepalingsmethode te construeren in overeenstemming met de Standards.

Daarom zal ik de Standards integraal doorfietsen en onderweg signaleren welke relevante aanwijzingen zij voor

cesuurbepalingsproblemen en hun uitwerking bevat.

Bij bespreking van dit uitgangspunt heb ik er al op gewezen dat deze in eerste instantie voor gestandaardiseerde tests en toetsen bedoelde Standards ook van toepassing zijn op het soort probleem waar het bij cesuurbepaling om gaat. Ook de docent die zelf zijn toets in elkaar zet, en op basis van scores op die toets uitspraken over studenten gaat doen en beslissingen over hen gaat nemen, heeft zich te houden aan bepaalde normen. Misschien ten overvloede: ook een mondeling tentamen dient aan deze normen getoetst te worden.

"These standards are written specifically to apply to standardized tests. They apply in varying degrees, however, to the entire range of assessment techniques."

In alle gevallen waarin cesuurbepaling plaatsvindt (mede) omdat men de beoordeling als voorspelling voor bijv. later studiesucces gebruikt, moet er voor gezorgd worden dat voor dergelijke criterium variabelen (in dit geval 'later studiesucces') behoorlijke metingen verricht worden (of zijn), en moeten dergelijke metingen met name ook aan dezelfde Standards voldoen.

"These standards also apply to criterion measures. A criterion measure should have the psychometric properties expected of any other measurement, such as validity, including in special instances some form of criterionrelated validity, for example, the relationship of an immediate criterion measure to an intermediate or more nearly ultimate measure. Criterion development should be guided by standards guiding test development."

In de praktijk betekent dit vooral dat de docent zichzelf niet moet toestaan bijvoorbeeld de cesuur te verscherpen omdat hij van mening is dat studenten met lage prestaties voor zijn vak in de verdere studie bijzondere moeilijkheden ondervinden. Een dergelijke opvatting hoort eerst getoetst te worden op zijn juistheid, een toetsing die alleen mogelijk is wanneer men in staat is zoiets als 'verder studiesucces' op een behoorlijke wijze meetbaar te maken. Ook dan kan onderzoek nog uitwijzen dat een eventueel bestaande correlatie niet hoeft te betekenen

dat er een oorzakelijk verband bestaat tussen slechte prestaties op een bepaald vak, en latere studiemoeilijkheden.

Zak-slaagbeslissingen worden helaas nog al eens genomen op grond van informatie over aanwezigheid op college, e.d.:

"A comment also seems appropriate about hidden tests (such as an interviewer's systematic attempt to assess a trait within the context of the interview) or other unobtrusive or observational measures. Some of these may raise ethical problems but they do not differ, in principle, from other tests, and the Standards apply as much to these unseen or unrecognized tests as to those more clearly perceived by the examinee."

Van bijzonder groot belang is de vraag welke gevolgtrekkingen gemaakt mogen worden op grond van toetsresultaten. Ofwel de vraag naar de validiteit van de toets voor de daarop te baseren beslissingen. Ik zal in het vervolg ongegeneerd over de validiteit van beslissingen, of de validiteit van de cesuur schrijven.

"Questions of validity are questions of what may properly be inferred from a test score; validity refers to the appropriateness of inferences from test scores or other forms of assessment."

De validiteit van beslissingen is alleen te evalueren door er [empirisch, bw] onderzoek naar te doen, door te onderzoeken of de resultaten van de beslissingen in overeenstemming zijn met de bedoelingen.

"There are various methods of validation, and all, in a fundamental sense, require a definition of what is to be inferred from the scores and data to show that there is an acceptable basis for such inferences."

Het kan dan ook niet voorkomen dat bepaalde toetsresultaten als 'voldoend en andere als 'onvoldoende' bestempeld worden zonder dat ooit een behoorlijke omschrijving van de voldoende of onvoldoende prestatie is gegeven; het spreekt vanzelf dat een struisvogel beleid dat de onvoldoende prestatie probeert te definiëren aan de toetsscore zelf, niet toelaatbaar is. Voor vrijwel iedere docent in het hoger onderwijs zal

het een bijzonder moeilijke opgave zijn om met een redelijk antwoord voor de dag te komen op de vraag wat een 'onvoldoende' bij hem betekent; en er is waarschijnlijk geen enkele docent die ook nog in staat is gegevens te produceren over de mate waarin zijn beleid bij het uitdelen

van onvoldoendes in overeenstemming is met zijn definitie van wat een onvoldoende is. Om precies in deze leemte te voorzien presenteer ik in dit rapport een alternatieve cesuurbepalingsmethode waarvoor wél geldt dat behoorlijk gedefinieerd is wat een onvoldoende prestatie is, en waarin de relatie tussen de definitie en de cesuurbepaling bekend is. (bij de kernitemmethoden en de methode Wijnen geldt dat zij niet valide zijn (de methode Wijnen) of dat over validiteit niets bekend is (de kernitemmethoden).

Zoals met veel zaken het geval is, komt validiteit ook in soorten. Het onderscheid is echter meer gemakshalve te maken, dan dat het hier zou gaan om strikt van elkaar onderscheiden varianten van validiteit. De Standaïda onderscheiden drie groepen validiteit: inhouds validiteit, predictive. validiteit, en construct validiteit.

"They are interrelated operationally and logically; only rarely is one of them alone important in a particular situation."

Voor cesuurbepaling in het onderwijs is om te beginnen de inhouds validiteit van het toetsinstrument van belang# Merk op dat de eis van inhouds validiteit in sterke mate parallel loopt met de eis van transparantie van de toetsing voor de student,

"To demonstrate the content validity of a set of test scores, one must show that the behaviors demonstrated in testing[^]constitute a representative sample of behaviors to be exhibited in a desired performance domain. Definitions of the performance domain, the users' objectives, and the method of sampling are critical to claims of content validity."

Het is wel duidelijk wat met inhoudsvaliditeit bedoeld wordt. Technieken voor het veilig stellen van inhoudsvaliditeit voor

studietoetsen kan ik hier niet bespreken. Een waarschuwing hoort er jammer genoeg nog wel bij: er is nogal eens verwarring tussen wat inhoudsvaliditeit is, en wat in het Amerikaans jargon aangeduid wordt met face validity. Het punt is van belang omdat met name de kernitemmethoden bijzonder vatbaar lijken voor verwarring van beide, ondanks alle goede bedoelingen van De Groot.

It should be clear that content validity is quite different from face validity. Content validity is determined by a set of operations, and one evaluates content validity by the thoroughness and care with which these operations have been constructed. In contrast, face validity is a judgment that the requirements of a test merely appear to be relevant. The writing of items in terms used in a particular job or by a particular subgroup of the population may give an appearance of relevance while contributing nothing to content validity or indeed to any other useful validity information.¹¹

Zolang het gaat om toetsen die door de docent in elkaar gezet worden, is het niet erg zinvol om construct validiteit te onderscheiden van predictieve validiteit. En predictieve validiteit, de naam zegt het al, heeft te maken met de vraag in hoeverre het terecht is dat op grond van scores op een toets verkregen, uitspraken worden gedaan die betrekking hebben op wat men verwacht dat scores voor een bepaalde student (of groep van studenten) op een andere (toekomstige) toets, examen e.d. zullen zijn.

Een kleine toelichting: bij cesuurbepaling hoeft het niet noodzakelijk zo te zijn dat kwesties van predictieve validiteit in het geding komen. Of dat het geval is, hangt af van de maatregelen die genomen worden op grond van gegeven voldoende of onvoldoende beoordelingen: wanneer van de student die een onvoldoende prestatie levert, verwacht wordt dat hij het studieonderdeel over zal doen, althans het tentamen nog een keer zal afleggen, wordt wél een voorspelling gedaan. De voorspelling is dat het voor deze studenten 'beter' is dat zij de stof nog een keer doornemen, dan dat zij doorgaan met het gewone studieprogramma; anders geformuleerd (ik heb er op diverse andere plaatsen in dit rapport al uitgebreid aandacht aan besteed) dat de student met een onvoldoende prestatie zijn tijd het 'best' kan

besteden door het nog maar eens te bestuderen, en niet door zijn tijd te besteden aan de studie van een ander vak in het programma. Cesuurbepaling in een onderwijs systeem waarin het voorgeschreven is om onvoldoende onderdelen alsnog voldoende te maken, moet dan ook kunnen voldoen aan de eisen van predictieve validiteit die hier in het geding zijn. Ik zal bij de argumentatie voor afschaffen pogingen om die predictieve validiteit te onderzoeken hoogst waarschijnlijk tot resultaat zullen hebben: de conclusie dat deze validiteit niet bestaat, of zelfs negatief is. (negatief: wanneer met het laten overdoen van studieonderdelen bedoeld wordt de student een betere uitgangspositie te verschaffen voor zijn verdere studie, en het zou blijken dat het gebruik van laten overdoen van studieonderdelen op zich de kansen op succesvol afsluiten van de studie verkleint, dan hebben we een verband gevonden dat tegengesteld is aan wat verwacht zou worden door voorstanders van het laten overdoen van onvoldoende vakken).

Het bovenstaande illustreert ook nog eens dat validiteit van een toets of een beslissing altijd staat in relatie tot de situatie waarin getoetst wordt, de oogmerken van de beslissingsprocedure e.d.:

"Statements about validity should refer to the validity of particular interpretations or of particular types of decisions."

Een passage in de Standards met speciale relevantie voor de kernitemmethode zoals voorgesteld door Van Naerssen (die nogal zwaar leunt op correlaties van items met totale toetsscore):

"Correlation of item scores with total scores on the test in which the item is included (or a parallel form of that test) may be presented as itemdiscrimination coefficients, but they should not be presented or used as itemvalidity coefficients. (Comment: itemdiscrimination coefficients are useful in reasoning about construct validity (* ...).However, they are indicators of internal consistency, not of validity.)"

Voor het verzamelen van gegevens over validiteit van een bepaalde beslissingsprocedure (methode voor cesuurbepaling in een bepaalde onderwijs context) geven de Standards een groot aantal

normen. Het uitvoeren van een valideringsonderzoek is zonder uitzondering een complexe aangelegenheid, en wel in het bijzonder omdat het behoorlijk opzetten, uitvoeren en interpreteren van dergelijk onderzoek vereist dat de onderzoeker in dit bijzondere probleemgebied goed is ingewerkt. Het moet de individuele docent of vakgroep dan ook met nadruk afgeraden worden om op eigen houtje dergelijk 'valideringsonderzoek' uit te voeren: men is daarvoor aangewezen op deskundige begeleiding zoals gegeven kan worden door onderwijskundige bureaus, of vakgroepen methodologie e.d. binnen psychologische faculteiten. Normen die bij dergelijk valideringsonderzoek in acht moeten worden genomen, zal ik hier dan ook niet bespreken. De docent die valideringsonderzoek laat uitvoeren, of onder deskundige begeleiding zelf uitvoert, blijft desalniettemin volledig verantwoordelijk voor de beslissingen die hij over studenten neemt, ook al meent hij daarbij op resultaten van valideringsonderzoek te kunnen steunen. Met andere woorden: de docent zal toch zélf moeten nagaan of zijn valideringsonderzoek voldoet aan normen zoals in de Standards gesteld. Het spreekt vanzelf dat universitaire folklore nooit ook maar enige evidentie voor wat voor validiteit dan ook kan zijn.

De betrouwbaarheid van beslissingsprocedures wil Ik hier maar heel kort behandelen. Je zou kunnen zeggen dat de betrouwbaarheid van een beslissingsprocedure gelijk is aan de predictieve validiteit van die beslissingsprocedure met betrekking tot een onafhankelijke replicatie van zichzelf. Dus. op dezelfde wijze als de docent gegevens moet hebben, verzamelen, of laten verzamelen over de validiteit van zijn beslissingsprocedure met betrekking tot de daarmee nagestreefde doelen, zo moet hij dat ook ten aanzien van de betrouwbaarheid van de beslissingsprocedure. Het is bijvoorbeeld mogelijk dat beslissingen beïnvloed worden door de persoon die de beslissing neemt (d.w.z. dat beslissingen anders kunnen uitvallen al naar gelang de ene of de andere docent het werk van de student beoordeelt). Het gaat er niet om dat onbetrouwbaarheid uitgebannen wordt, maar dat bekend is in welke mate de beslissingen onbetrouwbaar zijn, zodat eventueel de nodige maatregelen genomen kunnen worden. Een voorbeeld van zo'n 'maatregel' is de cesuurbepaling volgens Wijnen, waarin de cesuur lager uitvalt naarmate de standaard meetfout van de toetsscores groter

is. (De standaard meetfout is niet hetzelfde als de betrouwbaarheid, maar verdient in bepaalde gevallen wel de voorkeur:

"The standard error of measurement ordinarily is more useful; it has great stability across populations since it is relatively independent of range of talent, and it may be used to identify limits that have a defined probability of including the true score. Test users may use reliability coefficients in comparing tests, but they use standard errors of measurement in interpreting test scores. Information in a test manual about a standard error of measurement may often be more important than information about a reliability coefficient.").

In het algemeen geldt ook hier dat het onderzoeken van de betrouwbaarheid van een beslissingsprocedure of van een studietoets niet een opgave is waar lichtvaardig over gedacht kan worden. Met name is het betreurenswaardig dat het makkelijk kunnen berekenen van een homogeniteits-index als de bekende K.R. 20 leidt tot de misvatting dat hiermee ook een behoorlijke schatting van de betrouwbaarheid van de toetsscores verkregen zou zijn. Wat ik hier bedoel is onmiddellijk duidelijk te maken door een iets juistere formulering van de betrouwbaarheidsproblematiek te geven: een toets heeft niet één bepaalde betrouwbaarheid, omdat het er om gaat op welke wijze van de toetsscores gebruik gemaakt wordt. Het gebruik van de toetsscores moet gevalideerd worden, en zo moet ook de betrouwbaarheid van een toets bepaald worden naar het gebruik dat van die scores gemaakt gaat worden, of het soort interpretatie dat van de scores gemaakt wordt. Welnu, de berekening van K.R. 20 is een automatisme, dat de computer altijd wel wil uitvoeren, ook al heeft de docent niets maar dan ook helemaal niets van zijn bedoelingen met een bepaalde toets kenbaar gemaakt.

Er is een in principe oneindig aantal bronnen van onbetrouwbaarheid, namelijk al die factoren die op de scores van invloed zijn, behalve de bedoelde factoren zoals bijv. de mate van beheersing van de stof. Tenminste moet voor een bepaalde beslissingsprocedure aangegeven worden net welke storende invloeden men rekening wil houden, welke storende invloeden geacht worden niet van belang te zijn, en eventueel welke storende

invloeden waarschijnlijk wel aanwezig zijn maar op geen enkele wijze gecontroleerd kunnen worden. Stanley, in Thordike's Educational Measurement, geeft een uitstekend overzicht van bronnen van foutenvariantie, zoals dat wel genoemd wordt. Het is duidelijk dat betrouwbaarheidsonderzoek geen simpele zaak is, dat het een aanfluiting is om K.R. 20 waarden te gebruiken als indicatie van 'de' betrouwbaarheid van een toets, en dat men aangewezen is op deskundige bijstand wil men de nodige gegevens over betrouwbaarheid van een beslissingsprocedure verzamelen. (Voor het laatste zijn namelijk redelijk gecompliceerde experimentele designs nodig.) Het is niet gezegd dat je geen beslissingen zou kunnen nemen zolang niet alles over de relevante bronnen van onbetrouwbaarheid bekend is, maar het is wél zaak dat men zich realiseert op welke verschillende manieren de gegevens waar men zich bij die beslissingen op baseert, beïnvloed zijn door toevallige omstandigheden die eigenlijk op de beslissing geen enkele invloed zouden mogen hebben. Het laatste is vooral belangrijk bij al die beslissingen die over personen worden genomen, met vaak vergaande consequenties voor de betrokkenen,

Eisen van validiteit en betrouwbaarheid zijn niet het soort eis waaraan een toets of een beslissingsprocedure voldoet of niet: van iedere toets of beslissingsprocedure moet vanaf het allereerste stadium van ontwikkeling van de betreffende toets of procedure duidelijk zijn welke doelen er mee gediend moeten worden, welke informatie daarbij relevant is en waarvan dus zoveel mogelijk gebruik gemaakt moet worden, en welke informatie irrelevant is en dus zoveel mogelijk uitgeschakeld, gecontroleerd, of althans wat betreft mogelijke storende invloed bekend moet zijn. In overeenstemming daarmee wordt de toets, of de beslissingsprocedure ontwikkeld. In het ideal geval volgt na gereed komen van de toets een experimentele fase waarin uitgezocht wordt of het zaakje ook inderdaad de eigenschappen van validiteit en betrouwbaarheid áéeft waarop men bij de ontwikkeling gemikt heeft. Pas daarna wordt (na eventuele wijzigingen te hebben aangebracht op grond van die eerste evaluatie) de toets gebruikt om er ook in feite beslissingen op te baseren.

De doeleinden van de cesuurbepalingsmethode zoals ik die als

alternatief voor de huidige kernitemmethoden, methode Wijnen, en andere strikt chaotische vormen van cesuurbepaling heb geschetst, die doeleinden zijn heel eenvoudig, en zullen wat validering e.d. betreft geen grote moeilijkheden opleveren.

Het hoofdstuk 'Standards for the use of tests' bevat een aantal normen die geloof ik wel voldoende voor zich zelf spreken, zoals:

"A test user should have a general knowledge of measurement principles and of the limitations of test interpretations."

"A test user should know and understand the literature relevant to the tests he uses and the testing problems with which he deals."

"One who has responsibility for decisions about individuals or policies that are based on test results should have an understanding of psychological or educational measurement and of validation and other test research."

"The choice or development of tests, test batteries, or other assessment procedures should be based on clearly formulated goals and hypotheses."

"A test user should consider more than one variable for assessment and the assessment of any given variable by more than one method".

De laatste aanwijzing impliceert voor toetsing in het onderwijs dat men heel wat voorzichter moet zijn met eenzijdig gebruiken van bepaalde toetsingsvormen dan men gewend is.

"Test scores used for selection or other administrative decisions about an individual may not be useful for individual or program evaluation and vice versa."

"If specific cutting scores are to be used as a basis for decisions a test user should have a rationale, justification, or explanation of the cutting scores adopted." (comment:()) The intent is to recommend that test users avoid the practice of designating purely arbitrary cutting scores they

can neither explain nor defend)".

"If examinees are to be selected on the basis of a set of scores that displays different regression lines for use in predicting the same criterion in different subgroups of an applicant population, cutting scores should be established with great caution to avoid unfairness to members of one or more of the subgroups." (Comment: There are many defensible definitions of 'unfairness' in the literature, and techniques have been developed for setting cutting scores to minimize 'unfairness' as defined in these ways. Test users should keep abreast of the rapidly developing literature on this topic.)"

Het laatste punt is van belang in onderwijs situaties waar nogal uiteenlopende groepen studenten aan deelnemen. Te denken valt bijvoorbeeld aan studies waar studenten met sterk uiteenlopende vooropleidingen wettelijk aan deel . kunnen nemen. Andere te onderscheiden bijzondere groepen zijn vrouwelijke studenten (in die archaïsche onderwijs situaties waar ze in mondelinge tentamens min of meer openlijk gediscrimineerd worden, of althans de overtuiging hebben gediscrimineerd te worden; of studenten die zich onderscheiden door een bijzondere mate van stress gevoeligheid). Aan dit probleem van gelijkheid van kansen wil ik hier geen verdere aandacht schenken. In het rapport met de technische details van mijn cesuurbepalingsaanpak kom ik er nog uitgebreid op terug,

Praktische uitwerking van het principe dat iedere student die het onderwijs behoorlijk volgt een voldoende krijgt.

Zonder hier in details te gaan treden (dat zal gebeuren in een afzonderlijk rapport waarin de gedetailleerde methodiek beschreven zal worden) zal ik hier de maatregelen schetsen die in overweging genomen kunnen worden om het onderwijs in overeenstemming te brengen met het principe.

Ik mag er wel van uit gaan dat voor ieder vak in een studieprogramma een bepaalde tijd is uitgetrokken die de student geacht wordt er aan te besteden. Het is best mogelijk dat de 'programmatijd' is afgestemd op zoiets als een 'normstudent' of wat

voor andere rare constructies dan ook. Het is zelfs erg waarschijnlijk dat de voor een bepaald vak uitgetrokken tijd tenminste ruim genoeg geacht wordt om de student die uiteindelijk de studie zal afmaken in staat te stellen in één keer een voldoende op het afsluitende tentamen te scoren, maar te krap voor andere studenten (die het dan nog maar een keertje moeten overdoen). Ik ga hier niet in op de vraag in hoeverre dit soort veronderstelde verbanden regel zijn (het is makkelijk om aan te tonen dat er weinig van kan deugen); het is voldoende er op te wijzen dat hier sprake is van een studieprogrammering die bedoeld wordt een selectieve werking te hebben. Welnu, we hebben afgesproken cesuurbepaling en selectie uit elkaar te houden. Dat betekent bijvoorbeeld dat degenen die willen selecteren op grond van studieprestaties voor dit vak en andere vakken, dat kunnen doen door bijvoorbeeld een bepaald cijfergemiddelde over alle vakken als voorwaarde te stellen voor toelating tot een vervolg studie (doctoraal of research aantekening misschien). En dat betekent dat voor het bepalen van de cesuur we lare hebben aan verhalen over 'normstudenten' e.d.: het enige waar we hier mee te maken hebben zijn de studenten die toegelaten zijn tot deze studie(fase). Dan nog is het best mogelijk dat de docent vindt dat een bepaalde categorie studenten ,ten onrechte tot zijn onderwijs zijn toegelaten; in dat geval moet het probleem tenminste op (sub)faculteitsniveau besproken worden, en niet eigenmachtig, door de docent aangepakt worden door 'strengere' beoordeling,

Overleg op subfaculteitsniveau zou dan bijvoorbeeld kunnen resulteren in scherpere toelatingsselectie (als daartoe de mogelijkheden bestaan en een dergelijke scherpere selectie inderdaad valide genoeg is) of het ter beschikking stellen van meer tijd voor het onderhavige vak binnen het studieprogramma, of het reëel bijstellen van de kennelijk te hoge eisen voor dit vak, of een verbetering van het onderwijs wanneer daar knelpunten in blijken te zitten, e.d.

Dit gezegd zijnde, kan iedereen het wel eens zijn met het principe dat iedere tot het onderwijs toegelaten student bij een behoorlijke studie-inspanning in één keer voor het eindtentamen moet kunnen slagen. Het volgende probleem is dan of je zomaar iedereen waarvan je weet dat ze voldoende aandacht aan het vak besteed hebben,

een voldoende kunt geven. Natuurlijk kan dat, en natuurlijk moet dat. Daarnaast is het best mogelijk om door andere maatregelen de studenten die om wat voor reden dan ook prestaties leveren die een eind boven dat minimum niveau uitsteken, te 'belonen'. Het is dan ook geenszins zo dat het in principe iedereen maar een voldoende geven er in hoeft te resulteren dat voor bepaalde categorieën studenten een stuk (extrinsieke) studiemotivatie hierdoor wegvalt. Nog afgezien daarvan, dat in mijn voorstel de studiemotivatie voor een grote groep studenten die anders door de frustratie van het moeten overdoen van vakken het leven zuur gemaakt wordt, weer op peil gebracht wordt, geldt dat voor andere groepen studenten door andere maatregelen los van de cesuurbepaling de motivatie van buitenaf wat opgepoord kan worden. Inderdaad zal het vaak zo zijn dat een faculteit die in haar geheel gekozen heeft voor deze aanpak van de cesuurbepaling, ook een selectiebeleid geformuleerd zal hebben waarin bepaalde minimale gemiddelde prestaties gevraagd worden die uitgaan boven het gemiddeld alle vakken net voldoende gemaakt hebben: het is dan in het belang van de student om tenminste voor die vakken waartoe hij zich aangetrokken voelt prestaties te leveren die voldoende hoog zijn om eventuele slechtere resultaten op andere vakken te kunnen compenseren. Een nadeel van een dergelijk systeem kan zijn dat bepaalde vakken slecht gedaan worden door studenten die eerder in de studie al ruim voldoende punten wisten te verzamelen. Ook dán kan men maatregelen nemen om bepaalde ongewenste situaties te vermijden.

Overigens zou het een goede zaak zijn wanneer docenten wat meer onder druk gezet werden om hun onderwijs zo in elkaar te zetten dat ze er de belangstelling van de student mee trekken en vasthouden: een kwestie van er voor zorgen dat de intrinsieke motivatie gewekt wordt (nog altijd een belangrijk algemeen onderwijsdoel zou ik zeggen).

O.K., iedere student die zich behoorlijk inzet, krijgt die voldoende. Maar hoe weet je nu of die student zich inderdaad behoorlijk gedragen heeft in overeenstemming met wat van hem verwacht werd? Dit is inderdaad een vraag die gesteld mag worden, maar ik wil hem onmiddellijk aanvullen met de vraag hoe de docent er voor kan zorgen dat de student ook behoorlijk studeert. Dat hier een 'probleem ligt, is

duidelijk: er zullen altijd studenten zijn die de kantjes er af gelopen hebben, te weinig aandacht aan dit vak besteed hebben in vergelijking tot wat er in het studieprogramma voor gevraagd wordt. Voor een deel kunnen dat studenten zijn die door omstandigheden die in de persoonlijke levenssfeer liggen het even moeilijk gehad hebben met de studie; voor een deel zullen het mensen zijn die misschien terecht menen aan de minimum eisen wel te kunnen voldoen door er maar een klein deel van de programmatisch voor dit vak uitgetrokken tijd aan te besteden; en misschien houd je dan nog een enkele echte luierik over, hoewel ik er van overtuigd ben dat die categorie studenten erg klein is. Ik ga er dan maar even van uit dat de docent of de faculteit er in de onderwijsopzet voor gezorgd heeft dat de student niet uitgenodigd wordt tot het inefficiënt gebruiken van zijn studietijd.

Wat dat laatste betreft een extreem voorbeeld. Een faculteit die meent de student een jaar lang colleges te kunnen laten lopen en van hem verwacht dat hij zich gelijk ook voorbereid op de tentamens die alle tegelijk aan het eind van het jaar gehouden worden, een dergelijke faculteit verdient klappen. Op minder grote schaal komt vrijwel overal in het hoger onderwijs de misstand voor dat het onderwijs zo in elkaar zit dat de student niet anders kan dan inefficiënt studeren: bepaalde tentamens vallen in de tijd ongeveer samen; bepaalde zware tentamens moeten in één keer gedaan worden terwijl daar een voorbereidingstijd van meerdere maanden voor staat; op tentamens wordt een soort feitenkennis gevraagd die pas kort vóór het tentamen er behoorlijk ingestampt kan worden; de eisen voor een tentamen zijn zo onduidelijk dat je maar beter eerst een keer kunt proberen met een minimale voorbereiding; etcetera.

Het spreekt vanzelf dat het soort misstanden als hier opgesomd, binnen de kortste keren opgeheven moet worden wil er ook maar in de verste verte sprake kunnen zijn van een behoorlijke wijze van cesuurbepaling, en trouwens ook van behoorlijk en efficiënt onderwijs.. Voor een deel zijn de uitgangspunten ook gericht op het banen van de weg hiertoe. De eis van transparantie van de toetsingssituatie houdt vanzelf in dat de student in staat gesteld wordt zich zo kort mogelijk en efficiënt dus voor te bereiden op het tentamen en in het bijzonder op het behalen van tenminste een voldoende

daarop. Wanneer de docent constateert dat het merendeel van de studenten te kort voor het tentamen serieus met de studie begint, kan hij overwegen om door opsplitsing van het tentamen in meerdere (vrijstellende) onderdelen, of door het onderbrengen van een deel van de stof in Handelingsonderdelen in plaats van in de eindtoetsing, de onderwijsinspanning van de student beter over de onderwijsperiode te spreiden. (Handelingsonderdelen (een term geïntroduceerd door De Groot 1972): practicum onderdelen, werkgroepbijeenkomsten, scripties en andere concrete werkzaamheden, onderzoeken, e.d.; d.w.z. al die studieonderdelen waar het niet gaat om het leveren van een bepaalde prestatie, maar om het intensief ergens mee bezig geweest zijn, het gedaan hebben van onderzoek, praktijk, e.d.). Door dit soort verbeteringen in de onderwijs situatie aan te brengen probeert de docent te bewerkstelligen dat de student zijn studietijd behoorlijk en efficiënt besteedt. De andere weg, het proberen te volstaan met het controleren van de tijdbesteding van de student, is aanzienlijk moeilijker. en is eigenlijk ook tamelijk machteloos. Het is nog maar de vraag of bestaande methoden voor studietijdmeting hier gebruikt kunnen worden, wanneer de student weet dat wat hij invult tégen hem gebruikt kan worden. Het laten invullen van tijdbestedingsformulieren heeft bovendien geen donder met onderwijs te maken, en zou alleen al om die reden zo lang mogelijk vermeden kunnen worden. Wat eventueel nog wel het overwegen waard zou kunnen zijn, is om van tijd tot tijd een kleine steekproef te nemen, en zeg om de paar jaar een willekeurig gekozen groep studenten te vragen aan een tijdbestedingsonderzoek mee te doen dat ten doel heeft het gegeven onderwijs te evalueren.

De docent zal door een bepaalde opzet van onderwijs de student er zoveel mogelijk toe brengen (hem helpen zo je wilt) om zijn tijd effectief te besteden. In de praktijk is het dan ook voldoende, wanneer de docent over aanwijzingen beschikt dat de opzet voor de grote groep van studenten inderdaad slaagt, op grond daarvan er van uit te gaan dat iedere student een behoorlijke inspanning voor zijn vak geleverd heeft, tenzij het tegendeel bewezen kan worden. Wat ik hier bedoel is dat de docent niet moet proberen om voor iedere student afzonderlijk uit te zoeken of hij of zij wel behoorlijk aan zijn vak gewerkt heeft, wanneer hij al weet dat het merendeel, of vrijwel alle studenten dat

inderdaad doet.

Bij al die maatregelen om de student te verleiden de voorgeschreven uurtjes ook inderdaad vol te maken, gaat het voorzover de cesuurbepaling betreft, om de groep studenten die wat meer moeite met het vak hebben. De docent moet dan ook onderscheid maken tussen deze studenten die met de nodige inspanning juist aan de eisen kan voldoen, en andere groepen studenten die door betere voorkennis of door hogere (andere) capaciteiten wat minder moeite met het vak hebben. Het kan in de praktijk erg moeilijk zijn om met het ongewapend oog een onderscheid te maken tussen de lapswanzende student die het inderdaad allemaal al wel begrepen heeft, en de zwakke broeder die er ook nog niet toe kan komen om een behoorlijke studie-inspanning te leveren. Wanneer de docent daar moeite mee heeft, laat hij dan proberen door vroegtijdige toetsing die groepen studenten van elkaar te scheiden. Een mogelijkheid is bijvoorbeeld om aan het begin van het stuk onderwijs een toets af te nemen op basis waarvan vrijstelling voor het vak verkregen kan worden wanneer de prestaties aan de minimum eisen voldoen (en wanneer studenten die daaraan voldoen geen behoefte hebben om hun kennis van dit vak te vergroten). Minder extreme maatregelen liggen misschien wat meer voor de hand, zoals de student in de gelegenheid stellen het eindtentamen af te leggen op het moment dat hij denkt daarvoor klaar te zijn (wat bij vormen van computer gestuurd onderwijs mogelijk kan zijn, maar in de meeste gevallen ook best zonder computer inschakeling gerealiseerd kan worden).

De docent of vakgroep die deze wijze van cesuurbepaling wil invoeren in een subfaculteit waarin overigens nog op middeleeuwse wijze wordt getentamineerd, komt voor een aantal bijzondere problemen te staan. Het risico zit er in dat studenten dit vak als een 'makkelijk' vak gaan beschouwen. Dat is allemaal nog niet zo erg, zolang het maar niet zo is dat ook de zwakkere student dat idee krijgt. Met andere woorden: juist in deze wat geïsoleerde situatie is het zaak het onderwijs zó op te zetten dat de 'zwakkerel' studenten inderdaad de in het studieprogramma voor dit vak uitgetrokken tijd er ook aan besteden. Voor de wat 'snellere' studenten geldt dat de vakgroep maar moet proberen voor hen het onderwijs voldoende interessant te maken

dat zij inderdaad een beheersing van de stof opdoen die een eind ligt boven het gestelde minimum niveau voor de cesuur. (Hoewel zelfs voor de figuur van de éne vakgroep die op deze wijze van cesuurbepaling overgaat zal gelden dat bij examenbeoordelingen dan toch hoge prestaties voor dit vak in de overwegingen mede betrokken zullen worden, zodat wat het halen van hogere cijfers betreft de student best gemotiveerd zal blijven.)

Het uitgangspunt dat toetsing in het onderwijs transparant moet zijn doet misschien wat overbodig aan na uitgangspunt dat in principe iedere student in één keer moet kunnen slagen. Immers, wil aan dit uitgangspunt voldaan kunnen worden, dan zal in redelijke mate de toetsing inderdaad transparant moeten zijn. Het ligt zelfs voor de hand om te stellen dat binnen de grens zoals bepaald door het uitgangspunt dat in principe iedere student in één keer slaagt, de efficiëntie van dat onderwijs verhoogd wordt naarmate de toetsing transparanter wordt gemaakt. Met 'efficiëntie' is dan zoiets bedoeld als de mate waarin de inhoudelijke onderwijsdoelstellingen gerealiseerd worden. Het omgekeerde, dat het transparanter maken van de toetsingssituatie zou bewerkstelligen dat méér studenten dan in één keer slagen, gaat zeker niet op (tenzij expliciet maatregelen genomen worden om het percentage afwijzingen ook te verkleinen bij het vergroten van de transparantie). Het is belangrijk zich te realiseren dat perfect transparante toetsingssituaties heel best reuze selectief kunnen zijn (bijvoorbeeld kan dat met opzet het geval zijn bij vergelijkende examens voor toelating tot een studierichting waarin slechts een beperkt aantal studieplaatsen beschikbaar zijn).

Na het inventariseren van alternatieve mogelijkheden voor het opzetten van cesuurbepalingsprocedures daaruit een beredeneerde keuze doen.

Allereerst, het formuleren van een minimum eisen pakket, zoals dat ook aan de orde is geweest in de uiteenzetting over het zorgen voor transparantie van de toetsing, hoeft geenszins noodzakelijk één enkel mogelijk eisen pakket tot gevolg te hebben. Neem voor het vak dat gedoceerd wordt, het te gebruiken tekstboek in gedachten, en stel je voor dat daaruit een minimum eisen pakket afgeleid moet worden.

Het is duidelijk dat die minimum eisen op veel verschillende manieren uit dat tekstboek gekozen kunnen worden. Niet alleen kunnen geheel andere opsommingen van onderwerpen worden gemaakt, maar vooral ook kunnen een aantal varianten in mate van gevraagde beheersing beproefd worden: het minimum niveau zou bijvoorbeeld gedefinieerd kunnen worden als het tenminste éénmaal doorgelezen hebben van het hele boek, terwijl een radicaal andere formulering is tenminste een bepaald deel van het boek diepgaand bestudeerd te hebben. Het zal onder andere van de doelstellingen van onderwijs afhangen, welk soort minimum eisen formulering gekozen wordt. Daarnaast zijn de randvoorwaarden van beperkt beschikbare studietijd, beperkte voorkennis van studenten e.d., van belang. Het is misschien mogelijk om een soort optimaal minimum eisen pakket af te leiden (met behulp van dynamic programming bijv., zie Page e.a. 1976), of een aantal op zich ongeveer even goede, maar qua karakter nogal uiteenlopende eisen formuleringen. Het belang van een en ander voor het cesuurprobleem zou ook nog hierin kunnen liggen, dat bepaalde formuleringen leiden tot minder spreiding in studieresultaten van bepaalde groepen studenten, dan andere. Afhankelijk wederom van de onderwijsdoelstellingen die men heeft, wordt bij voorkeur gekozen voor een zo klein mogelijke spreiding in studieprestaties, dan wel voor een wat verder uit elkaar trekken van studieprestaties. Een meer gedetailleerde uitwerking zal ik hier niet van geven.

Een ander voor de hand liggend alternatief is om in plaats van het tentamen de beoordeling op andere gegevens te baseren. In bepaalde gevallen ligt het voor de hand om niet af te gaan op het leveren van intellectuele prestaties op een tentamen aan het eind van de rit, maar in werkstukken, discussies, en andere activiteiten tijdens het onderwijs. Voor een belangrijk deel gebeurt dat natuurlijk al in practica, bij het maken van scripties, het doen van eigen onderzoek e.d., maar het principe laat zich uitbreiden tot méér onderwijs onderdelen. (Een verschuiving dus naar Handelingsonderdelen, in terminologie van De Groot). Een afzonderlijke argumentatie voor het bovenstaande is dat beoordelingen die frequent plaats vinden tijdens het onderwijs' in intensief contact met de student, een tamelijk grote mate van betrouwbaarheid en validiteit kunnen krijgen wanneer ze tezamen genomen worden, een mate van betrouwbaarheid en validiteit waar

een afsluitend tentamen waarschijnlijk maar moeilijk tegen op kan. (betrouwbaarheid en validiteit vanzelfsprekend gerelateerd aan de doelen die men met cesuurbepaling nastreeft).

Het. bovenstaande is misschien een iets extreem alternatief, maar er zijn natuurlijk mengvormen mogelijk, waarbij informatie verkregen uit beoordelingen tijdens het onderwijs van werkstukken e.d. bijvoorbeeld, gecombineerd worden met de prestaties geleverd op een afsluitende toets (of een aantal deelttoetsen die al eerder zijn afgenomen). Dat dit een beetje ingewikkeld wordt, geef ik direct toe, maar zolang die meerdere ingewikkeldheid goed gemaakt wordt door hogere kwaliteit van de cesuur beslissingen, is ze verantwoord. Bovendien is het een soort ingewikkeldheid die zich prima in beleid laat organiseren, en vergelijk dat eens met de rituelen waarmee traditionele cesuurbepalings zittingen gepaard gaan.

Tja, dan is een vraag die ook nog eens beantwoord moet worden, of bij de cesuurbepaling rekening gehouden kan, mag of moet worden met eerdere studieprestaties van de student op andere vakken. Iedereen kent wel de figuur van de docent die een mondeling tentamen zit af te nemen met voor zich op tafel de kaart van de student waarop zijn studiegeschiedenis in cijfers staat afgebeeld. Je zou kunnen zeggen dat het oordeel van de docent door deze extra informatie op ontoelaatbare wijze beïnvloed wordt. En dat is in zekere zin natuurlijk ook zo. Aan de andere kant is het ook evident dat beslissingen over studenten beter genomen kunnen worden naarmate over meer informatie over de student beschikt kan worden# Dat zouden dan niet alleen voorafgaande studieprestaties kunnen zijn, maar bijvoorbeeld ook de onafhankelijk van de op het tentamen door de student behaalde prestaties, een door de docent geformuleerde verwachting over de prestaties van afzonderlijke studenten, gebaseerd op zijn ervaringen met de student in de onderwijs situatie. Deze ideeën zien er enigszins chaotisch uit, maar ze kunnen heel nauwkeurige betekenis gegeven worden, en met name zijn er scherpe regels over de wijze waarop dergelijke apriori informatie gecombineerd kan worden met nieuw verkregen informatie om tot de beste schatting te komen van de mate waarin deze student de stof beheerst (of hij voldoet aan de minimum eisen of niet). Het zou te ver gaan om van iedere docent

een dergelijke gecompliceerde aanpak van zijn cesuurprobleem te verwachten. Desalniettemin is het redelijk om te veronderstellen dat in een niet te verre toekomst procedures, hulpmiddelen, en begeleiding aanwezig zullen zijn die gebruikmaking van deze technieken praktisch voordelig maakt. (Voor toepassingen en principes zie vooral Novick & Jackson 1974).

Nog even doorgaand met alternatieve procedures# Het ligt ook erg voor de hand om de cesuur niet meer te bepalen op die éne totaal score, maar het tentamen wat gevarieerder in elkaar te zetten, en scores op verschillende onderdelen van het tentamen als uitgangspunt te nemen voor de cesurbepaling. Dan is het ook eenvoudig om er meteen andere gegevens, zoals beoordelingen van werkstukken, essays, uitgebreide open vragen e.d., bij te betrekken. Dat alles min of meer indachtig het principe in de Standards geformuleerd, dat beslissingen over personen op een verscheidenheid aan informatie gebaseerd moeten worden, althans bij voorkeur niet op maar één wijze van toetsen over één bepaald soort opgaven.

Een groot aantal varianten in de beslissingsprocedure heeft te maken met wat er na de beslissing gebeurt, dus met wat de gevolgen voor de student (maar ook voor de staf, en eventuele anderen) zijn van een genomen beslissing (waarbij vooral de niet te vermijden categorie van foutieve beslissingen van belang is). Het maakt nogal enig verschil voor de beslissingsprocedure uit wanneer aan het geven van een onvoldoende tevens gekoppeld is dat een student zijn studietoelage verspeelt, twee jaar moet wachten voor een nieuwe examengelegenheid, of na een aantal jaren studie alsnog gedwongen wordt (zich gedwongen ziet) de studie te staken, vergeleken met de situatie waarin een onvoldoende voor een vak gewoon als zodanig genoteerd wordt, of eventueel in een herkansing de volgende dag gerepareerd kan worden. Daartussen liggen dan nog weer vele mogelijkheden. De gevolgen verbonden aan bepaalde beslissingen worden overigens expliciet in de cesurbepaling betrokken wanneer bij een bepaalde gekozen procedure, de beslissing genomen moet worden waar dan precies de cesuur gelegd moet gaan worden

In het ideale geval kan de docent uit een min of meer naar

volledigheid strevende lijst van alternatieve mogelijkheden degenen kiezen die hij in overweging wil nemen, en dan vervolgens tussen dat beperktere aantal een keuze maken op grond van een gedetailleerde analyse, en eventuele doorrekening van consequenties van de alternatieven voor het onderwijs, de individuele studenten, etcetera.

Het spreekt vanzelf dat de keuze van cesuurbepalings aanpak uit een dergelijke alternatieven lijst niet op louter intuïtieve gronden hoort te gebeuren. De te hanteren methode moet passen bij de onderwijsdoelstellingen, moet 'betaalbaar' zijn, of misschien zelfs zo goedkoop mogelijk, moet voldoen aan een aantal restricties zoals bijvoorbeeld onze zes uitgangspunten. Het ziet er dan ook naar uit dat we hier te maken hebben met een behoorlijk complex beslissingsprobleem, tenzij zich de gelukkige situatie voordoet dat één alternatief evident met kop en schouders boven alle andere uitsteekt. Ook dan kan het geen kwaad dat de docent of de vakgroep haar intuïtieve voorkeur toetst aan een verdere analyse van de 'voorkeurs' methode, en die bijvoorbeeld vergelijkt met het daaropvolgend 'bestel alternatief.

De grootste moeilijkheid bij het kiezen uit het lijstje alternatieven zal wel zijn dat verschillende alternatieven op verschillende doelstellingen beter zijn dan andere, die weer goedkoper zijn of makkelijker te hanteren, politiek meer haalbaar, etcetera. De absurd simpele opvatting dat uitsluitend het studierendement van belang isg hoeft na bovenstaande opsomming niet nog verder ontkracht te worden.

Een illustratief lijstje doelstellingen van onderwijs die in het oog gehouden, respectievelijk geoptimaliseerd moeten worden:

- a. eenmaal tot het onderwijs toegelaten studenten moeten alle kans krijgen het studieprogramma binnen de daarvoor vastgestelde tijd te voltooien;
- b. bij het verdelen van de onderwijsinspanning wordt prioriteit gegeven aan de meest onderwijs behoeftige studenten;
- c. prestaties van studenten moeten aan bepaalde minimum niveaus voldoen;
- d, de motivatie van de student is voor een belangrijk deel voor de verantwoordelijkheid van de onderwijsinstelling;

e. naast inhoudelijke (cognitieve) zijn ook de affectieve doelstellingen van belang;

f. bij het streven naar zo goed mogelijk realiseren van de doelstellingen moeten affectieve, motivationele, en cognitieve doelstellingen met elkaar in 'evenwicht' blijven;

Het is waarschijnlijk dat altijd wel enkele doelstellingen uit zo'n lijstje met elkaar in 'conflict' zijn. Desalniettemin moet een beslissing toch maar genomen worden. Er zal dan ook veel van af hangen, op welke wijze de alternatieven gewaardeerd gaan worden naar de mate waarin 'de doelstellingen' gerealiseerd worden. Veel van de daarvoor beschikbare technieken vragen de docent om aan verschillende doelstellingen een relatieve waarde toe te kennen, en te schatten in hoeverre bepaalde methoden leiden tot realiseren van de verscheidenen doelen. In principe kan dan 'berekend' worden welke methode het beste resultaat geeft. Het probleem kan opgevat worden als een opgave voor Operations Research, of meer specifiek als thuishorend in het probleemgebied van Multiple Objective (or Criteria) Decision Making.

De belangrijkste complicatie is dat naast rekening houden met doelstellingen, ook kosten factoren niet uit het oog verloren mogen worden.

'Kosten' kunnen natuurlijk ook gezien worden als 'Opbrengsten' wanneer een ander gezichtspunt gekozen wordt. Bijvoorbeeld, stel dat een bepaald voorstel voor de cesuurbepaling er op neer zou komen dat door de bank genomen de studietijd met een half jaar gereduceerd wordt vergeleken met de traditionele aanpak van het cesuurprobleem. Het alternatief levert een voordeel van een half jaar voor iedere student. Anders gezegd: het traditionele systeem kost iedere student een half jaar. Ook hier kan weer een lijst van 'kosten' of 'opbrengsten' gemaakt worden, eventueel onder te verdelen in 'kosten e.d.' die zo gering mogelijk gehouden moeten worden, en 'kosten e.d.' die tenminste binnen bepaalde grenzen moeten blijven. Overigens zijn de uitwerkingsproblemen vrijwel hetzelfde als voor de doelstellingen lijst: de meeste kosten zijn niet zonder meer in guldens te kwantificeren, en moeten dan ook met behulp van speciale technieken althans met elkaar

vergelijkbaar gemaakt worden.

Zijn er, naast doelstellingen van onderwijs en kosten overwegingen, nog andere randvoorwaarden van belang, dan kunnen die door de docent aan de lijst nog toegevoegd worden.

Het is duidelijk dat bij dit alles van belang is dat redelijke inzichten bestaan in de mogelijke gevolgen die verschillende cesuurbepalingsvoorstellen kunnen hebben. Gevolgen, beschreven in termen van doelstellingen die al dan niet of slechts in bepaalde mate bereikt zijn, en kosten en opbrengsten die een bepaalde methode met zich mee brengt. Voor het behoorlijk kunnen inschatten van dergelijke gevolgen zou de docent (de vakgroep, eventuele andere beleidsinstanties) bij voorkeur over empirische gegevens (onder gecontroleerde omstandigheden verkregen) moeten kunnen beschikken. Voor nieuwe alternatieve methoden zijn dergelijke gegevens uit de aard der zaak niet beschikbaar, en dan zal een poging gedaan moeten worden om op basis van elders verkregen resultaten, of theoretische inzichten, in te schatten of overgaan op een dergelijke methode gewenst is. Vaak zal het, niet mogelijk zijn om bij wijze van experiment een bepaalde methode slechts bij een deel van de studenten toe te passen, omdat het risico altijd aanwezig is dat bepaalde studenten zich benadeeld zullen voelen omdat zij ofwel in de experimentele groep zitten, dan wel juist niet in de experimentele groep zaten.

Besliskundige benadering van de cesuurbepaling. (het e uitgangspunt).

Een besliskundige aanpak houdt in dat geprobeerd wordt de beslissingssituatie zo volledig mogelijk te analyseren, daarover de relevante gegevens te verzamelen, om er vervolgens langs rationele (wiskundige) weg de beste oplossing(en) voor te breken. Om de werkwijze te illustreren eerst een eenvoudige uitwerking van deze procedure:

Veronderstel dat een vakgroep, geconfronteerd met een groep

studenten van zeer uiteenlopende capaciteiten en vooropleiding, besloten heeft om de cursus die zij moet verzorgen in twee cursussen te splitsen: een langzame voor degenen met weinig relevante voorkennis, en een snelle of meer diepgaande cursus voor degenen met behoorlijke voorkennis. Hoewel de vakgroep de keuze tussen beide vormen van de cursus in principe aan de student zelf wil laten, wil zij toch de student een advies geven over welke vorm van de cursus hij of zij het 'best' zou kunnen volgen. Dat advies wordt gebaseerd op de resultaten van een korte voortoets (afge nomen voordat de cursus begint). Veronderstel dat op grond van een behoorlijk validatie onderzoek van die voortoets bekend is wat het verband is tussen scores op die voortoets en het al dan niet bevredigend doorlopen van de geavanceerde cursus, respectievelijk de langzame cursus. Om op dit moment onnodige subtiliteiten te vermijden veronderstel ik dat alleen van belang is of de student voldoet aan de minimum eisen voor het afsluitende tentamen voor de snelle resp. langzame cursus. (Met studenten die in de snelle cursus tot bijzonder goede resultaten komen hoeft .. "ook geen rekening gehouden te worden omdat deze categorie studenten nu juist geen beslissingsprobleem oplevert). De situatie kan geanalyseerd worden als een toelatings probleem voor de snelle cursus, met dien verstande dat iedere student die niet het advies voor de snelle cursus krijgt, geadviseerd wordt de langzame cursus te doen.

De beslissingen waartussen gekozen moet worden zijn dan tweerlei.

- a. adviseren tot de snelle cursus
- b. adviseren tot de langzame cursus.

Een advies dat opgevolgd wordt, kan tot resultaat hebben dat het advies

1. ten onrechte gegeven blijkt
2. terecht gegeven blijkt.

De combinatie van beslissingen en mogelijke uitkomsten levert de volgende vier varianten op:

- A. ten onrechte gegeven advies voor de langzame cursus

- B. terecht gegeven advies voor de snelle cursus
- C. terecht gegeven advies voor de langzame cursus
- D. ten onrechte gegeven advies voor de snelle cursus.

Aan ieder van deze mogelijke resultaten wordt een bepaalde waarde of utiliteit toegekend. De docent kan hier veel of weinig moeite doen om tot een goede toekenning van waarden te komen, dat hangt een beetje af van het belang van de beslissing, de ernst van de gevolgen, en de beschikbare technieken voor toekenning van waarden. Veronderstel dat in dit geval de beste uitkomst is een terecht gegeven advies om de snelle cursus te volgen: ken daaraan de waarde 1 toe. Veronderstel dat het slechtste resultaat is een advies om de snelle cursus te volgen dat ten onrechte gegeven blijkt: ken daaraan de waarde 0 toe. De beide andere uitkomsten krijgen dan een waardering tussen nul en één, laten we zeggen (zonder ons op deze plaats te verdiepen in technieken voor een goede handelwijze bij het bepalen van dergelijke waarden) dat dat oplevert als waarde voor een ten onrechte gegeven advies voor het volgen van de langzame cursus van .2; waarde van terecht gegeven advies voor volgen van de langzame cursus waarde .4.

Op grond van eerder uitgevoerd validerings onderzoek voor de te gebruiken voortoets is bekend wat de kans is dat iemand met een bepaalde score op die voortoets zal blijken te voldoen aan de minimum eisen van de snelle cursus: de kans op succes in de snelle cursus, horend bij een gegeven score X op de voortoets, noem ik P . Waar we naar zoeken, is de 'bestel cesuur' op de voortoets, dat is de score $X(\text{cesuur})$ die tenminste gehaald moet zijn om het advies voor het volgen van de snelle cursus te krijgen. Omdat uit validerings onderzoek het verband tussen scores X op de voortoets en slaagkans P voor de snelle cursus bekend is, zouden we het probleem opgelost hebben wanneer de waarde van P die hoort bij de 'bestel cesuur' X kunnen berekenen. En dat kunnen we nu

Laten we er van uit gaan dat studenten met dezelfde score X op de voortoets ook hetzelfde advies krijgen (alleen bij een beperkt aantal beschikbare plaatsen in één van beide cursussen zou er bijv. geloot kunnen worden tussen studenten met dezelfde voortoets score die echter niet allen tot die cursus toegelaten zouden kunnen worden).

Neem dan een bepaalde score X op de voortoets in gedachten, en stel de vraag of studenten met die score op de toets allen het advies voor de snelle, dan wel voor de langzame cursus zouden moeten krijgen.

a. adviseren tot de snelle cursus. De kans van slagen is bekend: P . Of anders gezegd: de verwachting is dat van deze groep studenten met score X de proportie P slaagt voor de snelle cursus. En de proportie $(1-P)$ zal er niet voor slagen. De verwachte waarde van deze beslissing is gelijk aan P maal de waarde van slagen voor de snelle cursus ($B=1$) plus $(1-P)$ maal de waarde van niet slagen voor de snelle cursus ($D=0$):

Als verwachte waarde van het advies voor de snelle cursus afgekort wordt tot $E(a)$:

$$E(a) = P(B) + (1-P)(D) = P(1) + (1-P)(0) = P$$

b. adviseren tot de langzame cursus. De verwachte waarde van deze beslissing:

$$E(b) = P(A) + (1-P)(C) = P(.2) + (1-P)(.4) = .4 - P(.2).$$

De beslissingsregel is eenvoudig: adviseer de snelle cursus als $E(a)$ groter is dan $E(b)$. Adviseer de langzame cursus als $E(b)$ groter is dan $E(a)$. En het zal nauwelijks een verrassing meer zijn te horen dat de cesuur precies daar ligt waar $E(a)$ gelijk is aan $E(b)$, met andere woorden op het punt waar het de vakgroep onverschillig blijft of studenten met die bepaalde score op de voortoets de snelle, dan wel de langzame cursus gaan doen.

De berekening van de cesuur eindigt dan ook met het oplossen van P uit:

$$E(a) = E(b)$$

$$P = .4 - P(.2)$$

$$P + P(.2) = .4$$

$$P(1.2) = .4$$

$$P = .4/1.2 = 1/3. \text{ [sic]}$$

Zoek op voor welke voortoets score X geldt dat de bijbehorende slaag kans P gelijk is aan $1/3$ (of daar het dichtst bij komt), en de cesuur is gevonden. In dit geval zouden studenten dus al gauw het advies krijgen de snelle cursus te gaan volgen.

Dit is in principe de hele methode. Cruciale punten daarin zijn een behoorlijke uiteenlegging van het beslissingsprobleem, toekennen van waarden of utiliteiten aan verschillende mogelijke uitkomsten contingent op de verschillende mogelijke beslissingen, gegevens over het verband tussen de kans op verschillende mogelijke uitkomsten en de scores op test of toets op basis waarvan de beslissingen genomen moeten worden, en een stukje rekenkunde om uit deze gegevens de cesuur te bepalen.

Het is een ondoenlijke zaak om deze techniek toe te passen bij de traditionele opvattingen over cesuurbepaling. De nodige gegevens zijn gewoon niet beschikbaar, het evalueren van mogelijke uitkomsten is een onmogelijke zaak omdat juist over de gevolgen van zak-slaagbeslissingen zo weinig bekend is. Als dit inderdaad zo is, dan zou alleen deze stand van zaken al voldoende argument zijn om onmiddellijk de traditionele wijze van nemen van zak-slaagbeslissingen te verlaten, omdat het hier gaat om pertinent irrationele maatregelen. De enige schijn van aanvaardbaarheid die traditionele cesuurbepalingsgewoonten hebben, is dat docenten zowel als studenten van haver tot gort vertrouwd zijn met deze irrationaliteit, en in zekere zin ook weten wat ze in deze aan elkaar hebben. Je zou kunnen zeggen dat er een soort afspraak bestaat tussen studenten en docenten wat betreft de wijze van cesuurbepaling waarover studenten het misschien niet helemaal met de docenten eens zijn, maar die wel een soort schijntransparantie verschaft. Schijn, natuurlijk, want je zou ook heel duidelijke cesuur regels kunnen opstellen gedefinieerd aan competitieve prestaties die duidelijk niets met de inhoud en doelstellingen van je onderwijs te maken hebben, en dan zou ook iedereen weten waar hij of zij aan toe is: voorbeelden zijn bijv. selectie op beoordelingen van gedrag, op resultaten van intelligentie tests, op sportiviteit etcetera.

Voorwaarde voor het ook maar kunnen beginnen met toepassing van een besliskundige benadering, is een tamelijk ingrijpende gewijzigde opvatting over doelen van de cesuurbepaling. Ik heb in de geformuleerde uitgangspunten een poging gedaan een goede basis voor een dergelijke gewijzigde opvatting te geven. Het spreekt vanzelf dat de richting waarin de verandering gezocht wordt, is naar betere expliciete overeenstemming met doelstellingen van onderwijs, bescherming van de persoon van de student, bewaking van de efficiëntie van het onderwijs.

In het cesuurbepalingsvoorstel dat ik in dit rapport doe, zijn de gevolgen voor de student van beslissingen rond de grens voldoende-onvoldoende zo klein mogelijk gehouden. Het zal in de praktijk een beetje afhangen van het beleid dat de onderwijs instelling (faculteit) in haar geheel voert, of niet toch langs achterdeurtjes weer sterke consequenties aan eventuele onvoldoendes worden gekoppeld (consequenties M.b.t. het kunnen doorgaan met vervolgstudies bijv.). Voorbeelden van beslist ongelukkige vormen van beleid zouden zijn om van de student die een onvoldoende behaalt voor een bepaald vak te eisen dat hij dat vak alsnog voldoende maakt, wanneer dat voor de student een ernstig verlies in, studietijd zou betekenen; of ook wanneer examenbeslissingen gekoppeld zouden worden aan het al dan niet behaald hebben van onvoldoendes voor bepaalde vakken. Zijn dit soort beleidsmaatregelen ongelukkigerwijs ergens van kracht, dan geldt wederom de situatie door De Groot (1964) beschreven, dat de docent niet kan overzien wat de gevolgen van zijn zak-slaagbeslissing voor de student kunnen zijn. Als de docent die gevolgen niet kan overzien, is hij niet in staat de besliskundige aanpak van cesuurbepalingsprobleem te volgen, en dat betekent dat hij teruggeworpen wordt op in laatste instantie altijd noodzakelijk irrationele procedures. (en dus ook onrechtvaardige procedures tegenover de betrokken studenten, en inefficiënte procedures m.b.t. economisch gebruik van de beschikbare middelen bij het zo ver mogelijk realiseren van de onderwijsdoelstellingen.). Maar dit lijkt allemaal teveel op een herhaling van dingen die al eerder en vaker aan de orde zijn geweest.

Nogmaals transparantie.

Twee soorten van transparantie zijn van belang: transparantie van de toetsing op zich (aard van de te verwachten vragen e.d.), en daarnaast transparantie van de beslissingsregels op grond waarvan bepaald wordt of een bepaalde prestatie als voldoende aangemerkt wordt of niet.

De transparantie eisen voor een beslissingsprocedure zijn wat moeilijker te formuleren dan de transparantie eisen voor de toetsing. Het is niet voldoende om, wanneer er ondubbelzinnige beslissingsregels vastgelegd zijn voor wat een voldoende en wat een onvoldoende prestatie is, die regels dan maar aan de student mee te delen. Met andere woorden: transparantie is nog heel wat anders dan objectiviteit. Misschien is het verstandig om dit transparantie probleem in tweegn te knippen, en van elkaar te onderscheiden de aanwijzingen die de student krijgt over de beste studiestrategie die bij of zij kan volgen om tenminste met bepaalde zekerheid een voldoende te behalen, en daarnaast informatie over de beslissingsprocedure die weliswaar op zich de student geen studie aanwijzingen verschaffen, maar hem wel zoveel inzicht in de gang van zaken geven dat hij zich 'eens' of 'oneens' kan verklaren met de wijze waarop de beslissingen genomen zullen worden. Het spreekt vanzelf dat het weinig zin heeft om de student die informatie te geven, wanneer hij vervolgens niet in de gelegenheid wordt gesteld om op basis daarvan de beslissingsprocedure te beïnvloeden (bijvoorbeeld door overleg tussen studenten en staf). Misschien is het nog beter om het begrip transparantie te beperken tot die informatie die de student kan gebruiken bij het bepalen van zijn studiestrategie, en het begrip aanvaardbaarheid voor de student te gebruiken voor zijn waardering van andere relevante informatie over de beslissingsprocedure. (De Groot 1972).

Onder die aanvaardbaarheid voor de student vallen dan ook moeilijke problemen als het accepteren van kleine verschillen die van belang kunnen zijn of je nog net een voldoende, dan wel een onvoldoende scoort. (wat De Groot rangschikt onder 'Billijkheidal problemen'). De student hoeft bepaalde (voorgestelde) maatregelen

natuurlijk niet zonder meer te slikken: het kan bijvoorbeeld best zijn dat naar de opvattingen van studenten de consequenties van het zakken voor een tentamen dermate groot zijn, dat ze niet af mogen hangen van het al dan niet bij toeval een item meer of minder fout gegokt te hebben. In dit voorbeeld zou gezocht kunnen worden naar alternatieven waarbij de gevolgen minder ernstig gemaakt worden, of waarbij de transparantie vergroot wordt (het vergroten van de betrouwbaarheid van de beslissingsprocedure zou ook overwogen kunnen worden, maar heeft waarschijnlijk weinig 'appeal' voor de studenten, en inderdaad het gestelde probleem blijft bestaan, zij het dat een minder groot aantal studenten er de 'dupe' van zal worden).

Informatie op grond waarvan de student de aanvaardbaarheid van de beslissingsprocedure kan beoordelen, is onder andere al die informatie die betrekking heeft op validiteit en betrouwbaarheid van de procedure in relatie tot de gestelde doelen. Ontbreekt dergelijke informatie, dan kan noch de student, noch de docent de waarde van welke cesuurbepalingsmethode ook naar behoren schatten.

Een andere overweging van aanvaardbaarheid is dat de student over voldoende informatie over de gevolgde procedure moet beschikken om te kunnen beoordelen of hij terecht of ten onrechte een bepaalde (voldoende of) onvoldoende heeft gekregen. De student moet over voldoende informatie kunnen beschikken om voor een commissie van beroep bijvoorbeeld te kunnen betogen dat de procedure aan bepaalde kwalificaties niet of onvoldoende voldoet. (Bijkomend punt: de student moet zijn eigen werk kunnen inzien, en de scoring of beoordeling kunnen controleren).

Nabeschouwing van de uitgangspunten: directe consequenties.

Hoe abstract de bespreking van deze uitgangspunten ook is, er vallen voor het directe onderwijs beleid wel enkele gevolgtrekkingen uit te maken, ook zonder dat deze nieuwe cesuurbepalingsaanpak op grond van de in het later te verschijnen 'schroefjes en moertjes' rapport tot in detail is uitgewerkt.

Als we de uitgangspunten stuk voor stuk nog eens langs lopen:

1. De docent of vakgroep kan nu al iedere verantwoordelijkheid voor selectie van studenten, bij een uitblijvend faculteitsbeleid ten aanzien van selectie, afwijzen. (Omdat die verantwoordelijkheid ook niet door de docent gedragen kan worden.) De (sub)faculteit kan nu al een beleid ten aanzien van examenregelingen opstellen waarin de ongelukkige koppeling met cesuurbepaling voor afzonderlijke vakken, in aanzienlijke mate doorbroken wordt.

2. De docent kan nu al beginnen met zijn toetsing (nog) transparanter te maken, door studenten meer informatie over de te verwachten vragen te verstrekken, en méér informatie die de student in staat stelt behoorlijk op een voldoende beoordeling te 'mikken'.

De (sub) faculteit kan vakgroepen aansporen om minder geheimzinnig te doen over de aard van de vragen in de eindtoetsing. Het vrijgeven van oude tentamenvragen kan bevorderd worden.

3. Toetsing aan de Standards is een minder eenvoudige zaak. Zolang echter nog niets of vrijwel niets over betrouwbaarheid en validiteit van zak-slaagbeslissingen bekend is, is de 'bestel mogelijkheid voorlopig gebruik te maken van de methode Wijnen, tenzij afsluitende tentamens eenvoudig te vervangen zijn door H(andelings) onderdelen.

4. In principe krijgt iedere student een voldoende. Realisering hiervan eist waarschijnlijk enige onderwijs verandering, expliciete doordenking van beoordelingsmethoden e.d. Zolang een en ander nog niet is afgerond, geldt ook hier dat hantering van de methode Wijnen niet direct in strijd komt met het principe. Ieder afwijzingspercentage dat behoorlijk boven dat wat bij de methode Wijnen verkregen zou zijn, uitkomt, is verdacht: d.w.z. dat ernstig rekening gehouden moet worden met de mogelijkheid dat de betrokken vakgroep verzuimd heeft onderwijs te geven.

.5 en 6. Met de uitwerking hiervan kan voor een deel ook al vast begonnen worden, voor een deel moet nog gewacht worden op het te volgen technische rapport, in deze docentencursus uit te brengen.

Argumenten voor afschaffen van het overdoen.

Bij de bespreking van de cesuurbepaling zijn lang niet alle in de eerste bladzijden van dit rapport genoemde argumenten voor afschaffen van het overdoen aan de orde geweest. Dat wordt nu ingehaald. Voordat afzonderlijke argumenten toegelicht gaan worden, is het prettig dat de lezer een behoorlijk inzicht heeft in de situatie xxx rond dat gebruik van over laten doen, en dan vooral de verschillende te onderscheiden groepen van studenten die in dat rollenspel meedoen. De groep studenten moet voor een behoorlijke analyse opgesplitst worden in een aantal deelgroepen, waarvan sommige bij verdere bespreking buiten beschouwing kunnen blijven, en andere in het licht van de schijnwerpers zullen komen te staan. Een voorlopige opsplitsing kan gebeuren langs de lijnen zoals in figuur 1. aangegeven.

1e tentamengelegenheid.

a	voldoende beoordeling	b	onvoldoende beoordeling	c	niet opgekomen
---	--------------------------	---	----------------------------	---	-------------------

aa	ab	ba	bb
terecht	onterecht	terecht	onterecht

[terecht onvoldoende verder uitgesplitst:]

baa		bab
te weinig tijd aan de voorbe- reiding besteed; waagde een gokje; wilde eens kijken welk soort vragen gesteld werden; kon zich door persoonlijke om- standigheden, of andere studie- verplichtingen onvoldoende voorbereiden.		heeft de in het studieprogramma daarvoor vastgestelde tijd wel besteed aan de voorbereiding op het tentamen, maar kwam desondanks tijd tekort voor een voldoende voorbereiding.

[bab verder uitgesplitst:]

baba	babb	babc	babd
maakt de studie af, zonder de onvoldoende ingehaald te hebben.	staakt de studie zonder de onvol- doende ingehaald te hebben.	staakt de studie na het tentamen over gedaan te hebben	maakt de studie af, na het tentamen alsnog voldoende ge- maakt te hebben

FIGUUR 1. Uitsplitsing van de deelgroepen waar het bij de problematiek van het overdoen van onvoldoende beoordeelde vakken om gaat. De deelgroepen waar we uiteindelijk mee te maken hebben,

zijn groepen A en B in de onderste rij.

[noot 2002: kies voor figuur niet-proportioneel lettertype]

Voordat de verschillende deelgroepen, en de redenen om ze al dan niet bij de verdere analyse mee te nemen, besproken worden, zal ik hetzelfde schema getalsmatig invullen. Stel dat 220 studenten het onderwijs volgen, en dat 200 studenten voor de eerste tentamengelegenheid opkomen. Een in het huidige stelsel niet onredelijke situatie zou dan zijn dat, zeg, studenten een voldoende krijgen (groep a), studenten een onvoldoende (groep b), en een lijstje voor alle deelgroepen in figuur 1 zou er zo uit kunnen zien:

deelnemers:	200
a. (voldoende)	150
aa (terecht)	125
ab (ten onrechte)	25
b. (onvoldoende)	50
ba (terecht)	30
baa (te korte voorbereiding)	20
bab (beschikbare tijd te kort)	10
baba (studeert af zonder overdoen)	1
babb (staakt studie zonder overdoen)	1
babd (staakt studie na overdoen)	4
babd (studeert af na overdoen)	4
bb (ten onrechte onvoldoende)	20

TABEL 1. Cijfervoorbeeld voor relatieve aantallen studenten in groepen.

[noot 2002: kies voor tabel niet-proportioneel lettertype]

[200 ==> 150 en 50; 150 ==> 125 en 25; 50 ==> 30 en 20 (A); 30 ==> 20 en 10; 10 ==> 1, 1, 4, 4 (B).]

De lezer kan zelf experimenteren met andere getallen. Daarbij moet hij dan wel een aantal gedragsregels in acht nemen: bepaalde getallen kun je niet zonder meer gaan wijzigen, omdat externe randvoorwaarden daarvoor in de weg staan, en bovendien kan geen enkel getal gewijzigd worden zonder op één of meer andere plaatsen in de tabel ook getallen te wijzigen.

Een voorbeeld van het eerste geval. Stel de docent vindt het aantal ten onrechte gegeven onvoldoendes een grove overschatting. Hij of zij kan dat getal alleen maar kleiner maken wanneer de betrouwbaarheid

van de cesuurbepalingsmethode dat toelaat, Wat algemener gezegd: getalsmatige invulling van de tabel voor een bepaalde onderwijs situatie zal moeten berusten op behoorlijke empirische gegevens.

Een voorbeeld van het tweede geval: Stel dat besloten wordt om minder onvoldoendes te gaan uitreiken. De aantallen studenten in groepen A en B zullen dan in het algemeen verhoudingsgewijs niet gelijk blijven: de verwachting is dat bij het laten zakken van de aftestgrens het aantal studenten in groep A verhoudingsgewijs sneller daalt dan in groep B. Het soort restrictieve voorwaarden voor veranderen van de getallen in de tabel, zoals boven geïllustreerd, is niet direct inzichtelijk; desondanks zal ik op deze plek geen poging doen om een overzicht van dat soort restricties te geven.

Een aantal eenvoudige voorwaarden waaraan getalsmatige invulling in de tabel moet voldoen, zijn direct uit het schema af te lezen. Als we gemakshalve de naam van de deelgroepen gebruiken als symbool voor het aantal studenten in die deelgroep, dan zijn er de volgende getalsmatige beperkingen:

$a + b$ = aantal deelnemers aan het tentamen

$aa + ab = a$

$ba + bb = b$

$baa + bab = ba$

$baba + babb + babc + babd = bab$

Zo zal een verandering in het aantal in groep $babd$ repercussies hebben voor de in te vullen getallen in bab , ba , en b .

Het laatste soort sommetjes ziet er als erg triviaal uit. Schijn bedriegt hier echter weer eens: het gaat er om dat bij argument en tegenargument de discussiepartners niet vrij zijn om op willekeurige momenten met bepaalde tegenwerpingen te komen, die consequenties hebben die gemakshalve maar niet verder uitgedplozen worden. Wie, (ik noem maar een voorbeeld zou willen tegenwerpen dat in het algemeen het aantal studenten in groep B toch groter is dan in groep A, die moet dat nog maar eens waar zien te maken, omdat een dergelijke veronderstelling wel eens gevolgen zou kunnen hebben voor pakweg

de relatieve verhouding van a tot b, die ver bezijden de realiteit liggen.

Wanneer dat van pas komt, kunnen meer deelgroepen onderscheiden worden dan in figuur 1 gedaan is. Misschien heeft de lezer er behoefte aan om de groep ten onrechte voldoende beoordeelde studenten verder op te delen in degenen die op de bonnefooi tentamen kwamen doen (zeg groep aba) en degenen die zich conscientieus voorbereidden maar binnenede beschikbare tijd niet 'rond' konden komen (groep abb).

Tenslotte is het interessant om, voordat de draad van het betoog opgevat wordt, nog een aantal cijfermatige varianten op tabel 1 te bekijken. Varianten zijn bijvoorbeeld te construeren door:

1) de verhouding voldoende/onvoldoende beoordeling te veranderen, dat is a/b . (verleggen van de cesuur dus)

2) de verhouding voldoende beheersing/onvoldoende beheersing te veranderen, dat zijn de aantallen $(aa + bb)/(ab + ba)$.

3) de aantallen studenten die zich aan hun deel van het 'contract' hebben gehouden, versus de aantallen die dat niet hebben gedaan. Daarvoor hebben we het extra onderscheid aba en abb nodig:

aba: zij die zich te kort voorbereid hadden, maar toch een voldoende beoordeling kregen,

abb: zij die de voorgeschreven tijd aan de voorbereiding besteedden, maar daar niet genoeg aan hadden. (Dus: $aba = abb = ab$)

Dat wil zeggen dat het hier gaat om de verhouding $(aa + bb + abb + bab)/(aba + baa)$.

Zo men wil, kunnen ook degenen die zich door overmacht niet voldoende konden voorbereiden, apart onderscheiden worden.

4) veranderen van de verhouding van aantal studenten die het naar eigen oprechte overtuiging behoorlijk hadden voorbereid, versus aantal studenten die wisten dat ze hun zaakjes nog niet voldoende kenden (maar zich in die opvatting ook best vergist konden hebben, evenals onder de andere groep studenten zullen zitten die hun eigen capaciteiten overschatten). Daarbij kun je nog een groep 'onzekeren' onderscheiden, die in omvang gereduceerd zou kunnen worden door de toetsing en de cesuurbepalings procedure transparanter te maken.

5) Nog ingewikkelder varianten dan 4) zijn te bedenken, die de lezer kan experimenteren met variaties die voor zijn eigen onderwijssituatie relevant lijken.

Het is nu mogelijk om systematisch een aantal varianten af te gaan. Veronderstel dat 25% een onvoldoende beoordeling krijgt:

$$(1) a = 75 \text{ en } b = 25.$$

Veronderstel dat in feite 75% de stof voldoende beheerst:

$$(2) aa + bb = 75 \text{ en } ab + ba = 25$$

Uit veronderstelling 2) volgt dat tenminste 75% zich aan het 'contract' heeft gehouden. Het zou weinig reëel zijn te veronderstellen dat alle studenten uitkwamen met de ter beschikking gestelde tijd (tenzij daar bij de onderwijsopzet en verzorging heel expliciet naartoe gewerkt zou zijn). Veronderstel dat 5% tijd tekort kwam (daardoor in feite de stof niet voldoende beheerste):

$$(3) abb + bab = 5$$

Uit (2) en (3) volgt:

$$(4) aba + baa = 25 - 5 = 20$$

(opmerking: getallen stellen percentages over de hele groep deelnemende studenten voor. Ik ga er van uit dat slechts in speciale gevallen er een goede argumentatie op tafel gelegd kan worden om méér dan 25 % van de studenten een onvoldoende te geven. $a=75$ is dus een soort ondergrens voor het aantal gegeven voldoende's. Ik ga er ook van uit dat er tenminste altijd wel voor gezorgd moet en kan worden dat niet meer dan 25% van de deelnemers in feite de stof onvoldoende beheersen: ook $aa+bb=75$ is een ondergrens. Dan volgt tevens dat het aantal studenten dat aan de beschikbare tijd niet voldoende heeft om zich voor te bereiden, ligt tussen de 5% en de 20%; ik ga er van uit dat de groep die de stof in feite onvoldoende beheerst, en die maximaal

25% is, nooit in haar geheel zal bestaan uit studenten die luit hun tijd gelopen zijn'. Ofwel, de restrictie op deze deelgroep studenten is:

$$(5) 0 < (abb + bab) < (ab + ba).$$

Voor een cijfermatige invulling van het model op grond van deze veronderstellingen hebben we nog de vrijheid om één grootheid zelf te kiezen. Echter, in die keuze worden we beperkt door de betrouwbaarheid van de beslissingsprocedure, waaronder hier verstaan kan worden het percentage foutieve beslissingen:

$$(6) ab + bb$$

Nauwkeurige schattingen voor de betrouwbaarheid van de cesuurbepaling kunnen in specifieke situaties verkregen worden door bekende technieken te gebruiken (Hofstee 1970), of door te werken met veronderstellingen over de verdeling van 'feitelijke stofbeheersing' voor deze groep studenten (zie bijv. Alf & Dorfman). In dit rapport zal ik op deze en andere mogelijke technieken niet ingaan. Wat slechts van belang is, is een grove schatting van het percentage misclassificaties. De schattingen die ik hier ga hanteren zullen de lezer misschien verrassen. Ik wil hier al de waarschuwing geven dat je met dergelijke schattingen niet makkelijk pessimistisch genoeg kunt zijn. Wie er absoluut van' overtuigd is dat de door mij gehanteerde schattingen aan de te hoge kant zijn, nodig ik uit om met gebruikmaking van de aanpak besproken door Alf & Dorfman 1967 aan te tonen dat bij redelijke aannamen er lagere foutenpercentages uitrollen.

Gezien veronderstellingen (1) en (2) is het niet onredelijk om als percentage misclassificaties 20 te hanteren. (De 50',ol met de hoogste toets scores verdienen die voldoende wel, voor een klein aantal slechtste prestaties is een onvoldoende beoordeling ook wel terecht, de overige zijn 'twijfelgevallen', waarvan een deel de beoordeling heeft gekregen die ze verdiende, een ander deel de onjuiste beoordeling). Natuurlijk kan de lezer andere waarden kiezen en daarmee de zaak doorrekenen.

De veronderstellingen voor de ie variant samenvattend:

- (1) $a = 75$ (beoordelingen)
 (2) $aa + bb = 75$ (feitelijke beheersing)
 (3) $abb + bab = 5$
 (7) $ab + bb = 20$ (% foute beslissingen)

Om weinig informatieve extra varianten te vermijden, veronderstel ik ook voor alle varianten dat de kans op een voldoende beoordeling niet groter (of kleiner) is voor een student die met zijn voorbereiding binnen de beschikbare tijd niet klaar kon komen, dan voor de student die om wat voor andere redenen ook onvoldoende voorbereid aan het tentamen deelnam:

$$(8) aba/baa = abb/bab. \quad (\text{en dat is weer gelijk aan } ab/ba)$$

Met behulp van (1), (2), (3), (7), en (8) kunnen de percentages in de kolom 'variant 1' in tabel 2 op simpele wijze berekend worden:

- (9) $a = 75$ daaruit volgt
 (10) $aa + ab = 75$. Uit (2) en (10) volgt
 (11) $bb = ab$. Uit (7) en (11) volgt
 (12) $bb = ab = 10$

Vervolgens vinden we op eenvoudige wijze dat

$$aa = 65, \quad ba = 15, \quad (b = 25 = ba + bb), \text{ en}$$

$$b a b = 5 (b a / (a b + b a)) = 5 (15 / 25) = 3$$

(formule af te leiden uit

$$(3) \text{ en } (8)). \text{ En } abb = 2, aba = 10 - 2 = 8, baa = 20 - 8 = 12.$$

Variaties in veronderstelling (3) hebben geen invloed op andere percentages dan die voor abb , bab , aba , en baa . In de nog te bespreken varianten laten we veronderstelling 3 dan ook maar achterwege.

deelgroep	symbool	%in deelgroep
voldoende beoordeeld	a	75 (aangenomen)
terecht voldoende	aa	65
ten onrechte voldoende	ab	10
te kort voorbereid	aba	8
te weinig tijd beschikbaar	abb	2

onvoldoende beoordeeld	b	25
ten onrechte onvoldoende	bb	10
terecht onvoldoende	ba	15
te kort voorbereid	baa	12
te weinig tijd beschikbaar	bab	3

TABEL 2.Cijfermatige invulling voor variant 1.

Het aantal interessante varianten is niet erg groot. Voor a zal ik als verschillende waarden kiezen 75, 90, of 95%, voor het percentage, studenten dat in feite de stof voldoende beheerst ook 75, 90 of 95. Dat zijn bij elkaar 9 varianten, die in tabel 3 samengebracht zijn. Een kleine uiteenzetting met betrekking tot de veronderstelde aantallen foute beslissingen. Uit de veronderstellingen kunnen drie percentages berekend worden:

- a) het minimaal mogelijke percentage foute beslissingen,
- b) het maximaal mogelijke percentage (niet interessant)
- c) het verwachte percentage bij random toekennen van beoordelingen.

De redelijkerwijs bij toetsgebruik te verwachten percentages foute beslissingen liggen, laten we zeggen, ongeveer tussen a) en c) in.

variant	1	2	3	4	5	6	7	8	9
a = ..	75	90	95	75	90	95	75	90	95
aa + bb = ..	75	90	95	90	95	75	95	75	90
minimum ab+bb	0	0	0	15	5	20	20	15	5
random ab+bb	37,5	18	11,5	30	14	30	27,5	30	11,25
gekozen ab+bb	20	10	6	24	10	26	24	24	9
(a en b gekozen)									
aa (ter. vold.)	65	85	92	70,5	87,5	71	73	70,5	88
ab (tooe vold.)	10	5	3	4,5	2,5	24	2	19,5	7
ba (ter. onv.)	15	5	2	5,5	2,5	3	3	5,5	3
bb (t.o. onv.)	10	5	3	19,5	7,5	2	22	4,5	2

TABEL 3.Negen cijfermatige varianten. Percentages terecht en onterecht gegeven voldoende en onvoldoende.

(Korte toelichting:

het minimum aantal foute beslissingen is het verschil tussen het gegeven aantal voldoende, en het aantal studenten dat in feite een voldoende verdiende. het random (verwachte) aantal foutieve beslissingen wordt verkregen door de te vergeven voldoende te verloten. Dus bijvoorbeeld voor variant 1, zou (gemiddeld) een kwart van degenen die een voldoende verdient, een onvoldoende krijgen, en driekwart van degenen die een onvoldoende verdient zou een voldoende krijgen:

$$0,25 (75) + 0,75 (25) = 150/4 = 37,5 \%$$

Het maximale aantal foute beslissingen in variant 1 wordt verkregen door iedereen die een onvoldoende verdient, een voldoende te geven, zodat ook evenveel studenten die een voldoende verdienen een onvoldoende krijgen, samen 50%.

Opmerking: veel van deze getallen zien er bijzonder 'onverwacht' uit; dit is een verschijnsel dat zich in het algemeen voordoet bij 2x2 tabellen (zoals hier de classificatie al dan niet voldoende gekregen versus al dan niet voldoende verdiend) die door vastgelegde randtotalen (zoals onze voorwaarden (1) en (2) maar beperkt kunnen variëren. Zie bijvoorbeeld Bishop e.a. 1975.)

Waarom sommige deelgroepen bij de analyse van het overdoen buiten beschouwing kunnen blijven.

Het is duidelijk dat de groep studenten die de eerste keer terecht een voldoende beoordeling krijgt, bij het probleem 'overdoen' geen rol meer meespelen. Deelgroep aa blijft prettig buiten beschouwing. En zo zijn er nog een aantal deelgroepen die, om diverse redenen, voor deze analyse niet relevant zijn, of dat op een heel specifieke manier zijn.

Degenen die ten onrechte een voldoende hebben gekregen, zal ik niet bij de analyse betrekken, omdat zij het tentamen niet zullen overdoen. Wanneer iemand van mening is dat deze categorie te groot is, om er geen rekening mee te houden, begrijp ik daaruit

dat hij kennelijk de aftestgrens zou willen verhogen, om zodoende minder studenten ten onrechte een voldoende te geven. Laat dat dan ook maar gebeuren, maar niet dan nadat men zich er rekenschap van heeft gegeven dat het verminderen van het aantal ten onrechte gegeven voldoende door heil, verlagen van de aftestgrens alleen maar kan gebeuren (*ceteris paribus*) door meer studenten ten onrechte onvoldoendes te geven.

Een deelgroep die van groot belang is bij een behoorlijke analyse van de misstand van het overdoen, is juist die groep studenten die ten onrechte een onvoldoende heeft gekregen. Overdoen betekent voor hen vrijwel uitsluitend tijdverspilling, en mogelijk heeft dit overdoen nog andere vervelende repercussies, zowel wat persoonlijke en financiële omstandigheden betreft, als nadelige effecten op de normale studievoortgang. Deze deelgroep zal bij de bespreking in de volgende paragrafen dan ook scherp in het oog worden gehouden.

Een bepaalde groep studenten met onvoldoende scores zal ik b uiten beschouwing laten: dat zijn de studenten die met een te korte of in ieder geval onvoldoende voorbereiding aan het tentamen deelnemen en daar voor zakken (of slagen). Ik ga er zonder meer van uit dat met enkele eenvoudige maatregelen bereikt kan worden dat deze categorie studenten verwaarloosbaar klein wordt. Een alternatief beoordelingsstelsel zoals het Grade Point Average systeem (zie blz. 28) zou dit kunnen bewerkstelligen. Maar een nog eenvoudiger maatregel is om iedere student voor ieder vak slechts één tentamengelegenheid te geven. Deze maatregel ziet er erg cru uit, en is dat natuurlijk ook wel wanneer aan een aantal voorwaarden niet is voldaan:

- 1.) examenregelingen mogen dan geen bepalingen bevatten dat je voor tenminste een zeer groot deel van de vakken een voldoende behaald moet hebben (anders zou je door een ongelukkig toeval voor een examen zakken wanneer zo'n eenmalige tentamenkans in een onvoldoende resulteerde; maar ook is een dergelijke regeling een allereerste voorwaarde om de tentamen stress voor studenten te verminderen).

- 2.) Studieprestaties voor verschillende vakken kunnen elkaar verregaand compenseren. Examenisen zullen dan ook gesteld

kunnen worden in termen van een minimaal behaald cijfergemiddelde. (zie blz. 28).

3. De tentamensituatie moet voldoende transparant zijn (zie blz. 29 e.v., en blz. 49) om de student die mate van zekerheid te kunnen verschaffen die nodig is wanneer hij onder deze alternatieve regeling voor een tentamen wil opgaan. Deze eis van transparantie is overigens natuurlijk ook van het grootste belang.

Dit alternatieve systeem zal natuurlijk ook zijn 'ongelukjes' met zich meebrengen. Zolang dat uitzonderingen zijn, kunnen die gevallen altijd individueel bekeken worden, en gepaste regelingen daarvoor voorgesteld.

Het gedane voorstel komt er op neer dat we verlost zijn van de voortdurende ellende voor studenten zowel als docenten, zoals die vastzit aan het systeem van herhalingsgelegenheden: er zijn geen herhalingsgelegenheden meer nodig (nogmaals: behalve in een klein aantal uitzonderingsgevallen, meestal overmacht situaties waarin de student buiten zijn schuld verzeild is geraakt). En voor de komende analyse zal die hele groep studenten die om wat voor reden dan ook met een te korte voorbereiding voor een tentamen opkomt, buiten beschouwing worden gelaten. Voor zover zij een onvoldoende gekregen zouden hebben is dat de deelgroep baa (zie figuur 1).

Een op zich ook al om andere redenen gewenste wijziging in het beoordelingsstelsel maakt onze problemen dus al een stuk eenvoudiger. Hetzelfde geldt voor het loskoppelen van cesuurbepaling (beoordeling dus) en selectie (waarop in het eerste deel van dit rapport al uitgebreid is ingegaan). De individuele docent heeft niet de taak om te selecteren, en beschikt trouwens ook niet over de middelen om dat op maar enigszins verantwoorde wijze te kunnen doen. Een andere kant van de zaak is echter dat bij een kwantitatieve analyse van wat er gebeurt in het gebruikelijke laten overdoen van onvoldoende gemaakte studieonderdelen het van belang is om degenen die de studie uiteindelijk niet zullen afmaken, uit de analyse te houden. Dat kan, zolang de veronderstelling juist is dat het al dan niet slagen voor dit bepaalde studieonderdeel niet in causale relatie staat tot het staken van de studie (als dat wél zo zou zijn, dan is een kritisch

onderzoek van die onderwijs situatie het allereerste dat zou moeten gebeuren). Voorzover de groep studiestakers bestaat uit studenten die niet aan de minimum eisen konden voldoen, mag aangenomen worden dat die beslissing niet hangt op de prestaties voor één enkel vak (dat zou ook onbehoorlijk zijn, omdat toevallige factoren daar een te sterke invloed op hebben). Voorzover de studie gestaakt wordt om andere redenen dan te geringe studieprestaties, kunnen deze studenten ook beter buiten de analyse gehouden worden (bijv. omdat zij op hun laatste studieonderdeel waarschijnlijk enigszins a typische prestaties leveren). In figuur 1 werden deze deelgroepen aangeduid als baba, babb, en babe, voorzover zij behoorden tot de categorie die kampte met te weinig studiegelegenheid. Voor een zuivere vergelijking zou eenzelfde indeling gemaakt kunnen worden voor de groep bb, noem deze respectievelijk bb.a, bb.b, en bb.c. Dan is bb.d de groep studenten die ten onrechte een onvoldoende kreeg, en afstudeerde na het studieonderdeel overgedaan te hebben.

Dan blijft als belangrijke groep over: de studenten die voor de eerste tentamengelegenheid te weinig tijd ter beschikking hadden, dan ook terecht een onvoldoende kregen, en voordat zij afstudeerden dit studieonderdeel overdeden. Wat is een 'terechte' onvoldoende trouwens in dit verband: alleen maar dat zij in feite niet aan de op dat moment gestelde minimum eisen konden voldoen. Dat laat de vraag open, of die minimum eisen niet te streng gesteld waren. Daarover straks meer.

Samenvattend zijn voor de analyse van de overdoe problematiek de groep die ten onrechte een onvoldoende kreeg, en de groep die een onvoldoende kreeg omdat hen te weinig tijd gegund was, van belang, voorzover zij inderdaad afstudeerden na het overgedaan hebben van dit studieonderdeel

De soms belangrijke groep studenten die met in feite onvoldoende voorbereiding een gokje waagt door aan het tentamen deel te nemen, kan door eenvoudige maatregelen ontmoedigd worden om dat soort gokjes te wagen, en is diengevolge voor de verdere analyse niet van belang.

Misschien voelt de lezer zich wat ongemakkelijk met mijn

'kunstgreep' een belangrijke groep studenten, die tenslotte 'terecht' een onvoldoende kregen, bij de verdere argumentatie buiten beschouwing te houden. Wanneer het stelsel van 'overdoen' deze studenten er toe brengt in tweede instantie zich toch behoorlijk op de leerstof te werpen, dan zou dat als verdienste van het 'over laten doen' gerekend kunnen worden? Er zijn verschillende redenen waarom dat niet het geval is:

a) Zoals uit tabel 3. afgelezen kan worden, worden altijd het minst aantal foute beslissingen genomen wanneer het percentage te geven voldoende gelijk is aan het percentage studenten dat de stof in feite ook voldoende beheerst. (De optimale cesuur zal overigens in de meeste gevallen ergens anders gelegd moeten worden, omdat ten onrechte gegeven voldoende niet even ernstig zijn als ten onrechte geven voldoende. Zie blz. 46-48).

Zijn beide percentages aan elkaar gelijk, dan is in tabel 3. te zien dat naarmate een groter percentage 'in feite voldoende beheersing' in de groep deelnemende studenten bereikt kan worden, het aantal foute beslissingen zal afnemen: de varianten 1, 2, en 3 zullen in veel onderwijs situaties een aantal foute beslissingen opleveren van respectievelijk ongeveer 20%, 10% en 5% (wanneer percentage in feite voldoende stijgt van 75 naar 90 naar 95).

b) Voor zover deze groep bestaat uit studenten die door omstandigheden buiten hun macht zich onvoldoende konden voorbereiden, is het altijd beter voor hen direct meer passende maatregelen te nemen.

c) Voor zover deze groep bestaat uit studenten die te laat met de voorbereiding van het tentamen zijn begonnen, moeten onderwijsmaatregelen genomen worden die een meer regelmatige studie bevorderen (grote tentamens opsplitsen, vrijstellende deeltolsten, praktische (Handelings) onderdelen die gedeeltelijk vrijstellend kunnen zijn, bijtjdse diagnostische toetsingsgelegenheden, tentamengelegenheden niet meer op enkele, mogelijk slechts uitkomende momenten in het studiejaar, maar doorlopend (mogelijk bij Computer Managed toetsen bijv.) e.d.)

d) Voorzover er belangrijke aantallen studenten zijn die een

tentamen eerste willen 'proberen', moeten maatregelen genomen worden die de transparantie van toets en beslissingssituatie verhogen, die studenten die lijden onder 'examenvrees' een hen tegemoetkomende toetsstrategie aanbieden, die álle studenten in de gelegenheid stelt een 'proeftentamen' te komen doen etc.

e) Voorzover er nogal wat studenten zijn die een 'gok' wagen, kunnen eenvoudige maatregelen genomen worden die dat ontmoedigen: wanneer de transparantie voldoende is hoeft niemand meer te gokken, wanneer niet onnodig meerkeuze items gebruikt worden wordt het 'gokelement' in het tentamen zélf gereduceerd, wanneer de betrouwbaarheid van de toetsing opgevoerd wordt kan de student uitgelegd worden dat 'gokken' geen zin heeft, wanneer een alternatief beoordelingsstelsel wordt gehanteerd met éénmalige tentamengelegenheden en een variant van Grade Point Average systemen voor de uiteindelijke exameneisen is er geen plaats meer over voor 'gokken' (behalve voor een enkel studieonderdeel wanneer de student het gezien zijn andere prestaties wel meent te kunnen lijden).

b t/m e laten dus zien dat de groep studenten met een te korte voorbereiding op het tentamen door onderwijsverbeterende maatregelen vrijwel tot nul gereduceerd kan worden.

f) Het zal nooit mogelijk zijn om, zolang een groot aantal studenten het in feite tekort voorbereid heeft, vrijwel al deze studenten een onvoldoende te geven, zonder tegelijkertijd een zeer groot aantal studenten die het in feite wél voldoende beheerst een onvoldoende te geven.

g) Omgekeerd, zal een cesuur die de 'bonafide' studenten rechtvaardig behandelt, altijd een belangrijk aantal studenten een voldoende beoordeling geven die het in feite tekort hebben voorbereid.

h) Tussen.haakjes: als het gestelde onder f) en g) niet juist zou zijn, is het kennelijk mogelijk om degenen die het voldoende hebben voorbereid goed te onderscheiden van degenen die dat niet deden: in dat geval bestaat er binnen de kortste keren geen groep deelnemers meer die het tentamen tekort heeft voorbereid, omdat zij ervan verzekerd kunnen zijn geen voldoende te behalen. Dit punt 'tussen haakjes' omdat een dergelijke situatie dermate fictief is dat ze in de

praktijk vrijwel nooit te verwachten is.

Dus, of we het overdoen af willen schaffen of niet, in ieder geval zouden maatregelen genomen moeten worden om er voor te zorgen dat degenen die opkomen voor een tentamen de stof ook op voldoende niveau beheersen, met als enige te overwegen uitzondering de studenten die volgens het studieprogramma te weinig tijd hadden om zich voldoende voor te bereiden (de categorie bab).

Het omgekeerde hoeft dan nog niet noodzakelijk waar te zijn: dat voor het beslissen of overdoen gehandhaafd blijft of niet, het niet van belang is of de categorie studenten die met een te korte voorbereiding deelneemt te verwaarlozen is, of niet. Dit geval is echter ook eenvoudig:

a) deze groep is verwaarloosbaar klein, en dat is dan in overeenstemming met mijn uitgangspunt bij de volgende argumentatie om deze deelgroep in die beschouwing niet te betrekken; of

b) deze groep is niet te verwaarlozen wanneer overdoen gehandhaaft blijft, maar ook dan is het zo dat overgaan tot het nieuwe beleid om geen herhalingsgelegenheden (behoudens voor uitzonderingsgevallen) te geven tot gevolg zal hebben dat studenten niet meer lichtvaardig aan een tentamen zullen deelnemen. Houd de logica hiervan in de gaten: een belangrijk aantal 'niet bonafide' deelnemers onder de traditionele 'overdoen' praktijk is op zich een bijzonder nadelige situatie, óók binnen het beleid om onvoldoendes te laten overmaken, en kan dan ook niet als voordeel voor 'overdoen'-beleid worden aangevoerd. En anderzijds brengen de maatregelen nodig om overdoen af te schaffen met zich mee dat alleen bij uitzondering studenten onvoldoende voorbereid aan tentamens zullen deelnemen.

Een beetje langademig, dit betoog, maar het punt waar het om gaat is erg belangrijk, gezien de grote consequenties die de nu volgende argumentatie kan hebben, een argumentatie die steunt op de veronderstelling dat met de groep studenten die te kort voorbereid aan tentamens deelneemt, geen rekening hoeft worden gehouden bij de

afweging van voor en nadelen van het al dan niet laten overdoen van onvoldoende studieonderdelen.

Desondanks:

Stel dat het lukt met het soort maatregelen als geschetst op blz. 59 om het percentage deelnemers aan het tentamen dat de stof in feite voldoende beheerst op te voeren tot zeg 95%, dan is het ook evident dat de vraag of je het overdoen zou moeten afschaffen of niet, een groot deel van zijn praktische belang heeft verloren. Het gaat dan immers nog maar om een relatief heel klein aantal studenten. Dat geef ik dan ook onmiddellijk en grif toe. Maar ik wil daar toch de opmerking bij maken dat afschaffen van het overdoen op zich een beleidsmaatregel is waarmee een afname van het aantal 'niet bonafide' tentamen deelnemers te bereiken is. Ook dan zou afschaffen van het overdoen een ondersteuning van de andere beleidsmaatregelen zijn.

Overige argumenten voor afschaffen van het overdoen.,

Na voorgaande voorbereidende paragrafen, dan nu bespreking van de argumenten voor afschaffen van het overdoen, die nog niet bij de behandeling van de cesuurbepaling aan de orde zijn geweest.

Argument 1, Het is moeilijk of niet aantoonbaar dat studenten die zich de eerste keer behoorlijk voorbereidden, en een onvoldoende behaalden, de stof werkelijk beter beheersen wanneer zij de tweede keer wél een voldoende behalen.

De stelling zou nog uitgebreid kunnen worden: als betere stofbeheersing al aantoonbaar is, dan gaat het om een zo klein verschil voor een zo klein aantal studenten, dat deze 'winst' in geen redelijke verhouding kan staan tot de 'kosten' aan de operatie 'overdoen' verbonden.

Wat is een 'behoorlijke' voorbereiding? Dat is ófwel een voorbereiding binnen de tijd, of ongeveer in de tijd, in het

studieprogramma uitgetrokken voor dit studieonderdeel, die resulteert in een in feite voldoende beheersing van de stof; ofwel een voorbereiding die ongeveer de hele beschikbare tijd opslokte, maar desondanks niet resulteerde in een voldoende (tenminste minimale) feitelijke beheersing van de stof.

Om misverstanden meteen te voorkomen: met 'feitelijke' beheersing wordt niet de prestatie op de studietoets bedoeld. De toets is een instrument om die zogenaamde 'feitelijke' beheersing te meten, en die meting gebeurt noodzakelijkerwijs zeer gebrekkig. Er is in de praktijk altijd een fors verschil tussen de scores behaald op de toetsing, en de feitelijke beheersing van de stof, omdat die 'feitelijke' beheersing met een grote meetfout vastgesteld wordt.

'Feitelijke' beheersing is een platonisch begrip, zo je wilt, dat niet op zich waar te nemen is, maar dat' gepostuleerd kan worden. Hoewel feitelijke beheersing niet waarneembaar is, is vaak met enige moeite wel een uitspraak te doen hoe 'dicht in de buurt' de toets score ligt bij de 'feitelijke' beheersing. (De lezer die haring of kuit wil hebben van dit verhaal, dat er inderdaad verwarrend genoeg uitziet, kan zijn toevlucht nemen tot het tweede hoofdstuk van Lord & Novick 1958).

Teruggaand naar figuur 1 op bladzijde 51, bestaat de deelgroep die zich 'behoorlijk heeft voorbereid' uit de groepen aa, bb, en bab, *) zoals na de gegeven toelichting eenvoudig is in te zien. en daarvan kregen de groepen bb en bab de eerste keer een onvoldoende beoordeling.

Deze deelgroepen zijn voor de analyse alleen van belang voorzover de studenten inderdaad afstuderen na het alsnog voldoende gemaakt hebben van dit studieonderdeel, Ik heb verzuimd de redenen hiervoor in de voorgaande paragraaf behoorlijk aan te geven, daarom alsnog een stukje toelichting:

*) de groep die ten onrechte een voldoende krijgt, ab, zal voor een deel bestaan uit studenten die aan de beschikbare tijd niet genoeg hadden om zich voor te bereiden. Deze deelgroep, we noemen hem abb, werd in figuur 1 niet afzonderlijk aangegeven. Dit groepje is voor de argumentatie pro en contra overdoen van marginaal belang, omdat zij onder beide regelingen zou bestaan. Mogelijk dat al dan niet overdoen invloed heeft op

de verhouding in aantallen bab t.o.v. abb.

In sommige gevallen kan de student gebruik maken van een examenregeling die toestaat een of meer onvoldoendes te compenseren door betere studieprestaties voor andere vakken. Deze studenten studeren af zonder de onvoldoende alsnog in een voldoende omgezet te hebben. Het is duidelijk dat het voor deze studenten, althans voorzover het om dit bepaalde vak gaat, niet uitmaakt of het 'overdoen' als systeem geldt, of dat het 'overdoen' afgeschaft is. Deze studenten kunnen zowel in de deelgroep bb voorkomen, als in de deelgroep bab, en misschien kan in het vervolg nog gebruik gemaakt worden van de redelijke veronderstelling dat studenten voor wie deze omstandigheden gelden relatief meer in deelgroep bb te vinden zijn dan in deelgroep bab. Eerder heb ik de onderhavige categorieën studenten benoemd als bb.a en babb.

Evenzo is duidelijk dat voor afweging van pro's en contra's voor afschaffen van het overdoen deenen buiten beschouwing kunnen blijven die de studie staken (om welke redenen dan ook) zonder het betreffende studieonderdeel overgedaan te hebben. Dat zijn de categorieën bb.b en babb.

Voorzover dit staken van de studie niet veroorzaakt wordt door de behaalde onvoldoende, is het weglaten van deze categorieën studenten geen enkel probleem.

Voorzover studiestaking gemotiveerd is door behaalde slechte studieprestaties, ligt de zaak mogelijk iets ingewikkelder. Om te beginnen wil ik meteen de mogelijkheid uitsluiten dat studiestaking door een enkele onvoldoende afgedwongen kan zijn: dat is een misstand die niet aanvaardbaar is. Maar dan is de onderhavige onvoldoende wel een slecht studieresultaat in een overigens wat grotere serie van slechte studieresultaten.

Hoewel het stelsel van overdoen-op-grote-schaal ongetwijfeld een ongelukkige selectieve werking heeft, in de zin dat velen door het 'pech' hebben ten onrechte de conclusie trekken dat dit een verkeerde

studie voor hen is ('ten onrechte' = als zij door waren gegaan zouden zij behoorlijk afgestudeerd zijn), zal ik dit buiten de 'overdoen'-analyse laten, omdat selectie voor de studie als geheel ver uitgaat boven de directe gebeurtenissen rond het al of niet overdoen voor een enkel studieonderdeel waar ik het hier over heb,

Tenslotte wil ik voor behoud van de overzichtelijkheid ook maar buiten beschouwing houden degenen die de studie staken na dit studieonderdeel alsnog voldoende gemaakt te hebben. Het weglaten van deze categorieën bb.c en babc is alleen maar in het voordeel van verdedigers van 'overdoen', hetgeen voor deze studenten vrijwel louter (in ieder geval omdat het voor de groep bb.c) een verspilling van tijd en energie was. Het is dan ook niet uitgesloten dat ik later alsnog deze roep studenten in de analyse zal betrekken. Maar voorlopig maar liever niet dus.

Blijven over voor onze analyse de deelgroepen bb.d en babd, de studenten die afstuderen na hun onvoldoende (terecht of ten onrechte gekregen resp.) ingehaald te hebben. Op zich is dit al een heel belangrijk resultaat: in de meeste gevallen mag namelijk verwacht worden dat dit nogal kleine groepjes zijn, en bovendien mag verwacht worden dat groep bb.d omvangrijker is dan groep babd d.w.z. dat méér studenten het tentamen ten onrechte moeten overdoen dan die het tentamen over moeten doen omdat de studielast voor dit onderdeel niet behoorlijk op hun capaciteiten, voorkennis e.d. was afgestemd.

Opmerkelijk is dat in het dagelijkse denken over afleggen van tentamens en het geven van herhalingsgelegenheden, vrijwel niemand zich schijnt te realiseren hoe groot de groep studenten is die het tentamen overdoet omdat zij ten onrechte de eerste keer een onvoldoende kreeg. De docent die daar een indruk van wil krijgen, zou op basis van de volgende aanpak een schatting kunnen maken (zie voor de details Hofstee 1970):

Neem het laatst afgenomen tentamen, en verdeel de vragen zo goed mogelijk in twee groepen, zodat de ene groep vragen een ongeveer even goede afspiegeling van de tentameneisen is als de andere groep vragen. Geef iedere student op basis van beide groepen vragen

afzonderlijk een beoordeling: voldoende onvoldoende op dezelfde wijze als dat eerder voor alle vragen bij elkaar gebeurd is. Tel dan het aantal studenten die volgens de beoordeling op ieder van de 'helften' beide keren een voldoende, beide keren een onvoldoende, en beide keren een andere beoordeling (voldoende en onvoldoende, of omgekeerd) kregen. Dat is een grove benadering van het aantal fouten dat bij de zak-slaagbeslissing gemaakt wordt. Ik zeg met opzet een 'grove', benadering, omdat o.a. geen rekening is gehouden met het feit dat ieder van deze beide tentamens natuurlijk maar half zo lang is als het werkelijk gebruikte tentamen (dat betekent dat voor het hele tentamen het aantal fouten iets minder zou zijn), en ook geen rekening is gehouden met de mogelijkheid dat het afwijzingspercentage afwijkt van het percentage dat in feite de stof voldoende beheerst (hoe groter die afwijking is, des te meer fouten worden er gemaakt boven het aantal dat op de aangegeven wijze geschat wordt).

In tabel 3. werd een schatting gemaakt van het aantal foute beslissingen zoals dat in de praktijk ongeveer gemaakt zal worden. ,

Het is erg belangrijk zich het bestaan te realiseren van deze groep studenten die iets zit over te doen wat ze in feite de eerste keer ook al behoorlijk beheersten. Spreekt men over de zinvolheid van het laten inhalen van 'onvoldoendes', dan moet vooral deze groep bij de afweging 'worden betrokken, omdat voor hen nu juist geldt dat de hele exercitie bij uitstek zinloos is.

Ja maar, zou daar tegenin gebracht kunnen worden, waarschijnlijk zullen deze studenten de tweede keer dan toch hun zaakjes wel iets beter beheersen, al was het alleen maar omdat zij nu minder risico willen lopen ten onrechte afgewezen te worden (gesteld al dat zij zelf er ook van overtuigd waren ten onrechte een onvoldoende te hebben gekregen). Is dat wel zo? Zou je volgens deze redenering er dan ook niet voor moeten pleiten om dan iedereen, ook degenen die de eerste keer al een voldoende hadden, het tentamen nog maar eens over te laten doen? Dat zou een absurde zaak zijn, maar desalniettemin, als men redeneert dat het tenminste geen kwaad kan om studenten die ten onrechte een onvoldoende kregen het zaakje nog maar eens te laten overdoen, dan vereist de logische coherentie in opvattingen dat

dat dan ook voor alle andere studenten geldt. Terug redenerend valt alleen maar te concluderen, hoe je het ook bekijkt, dat voor studenten die ten onrechte een onvoldoende kregen het overdoen uitsluitend een verliesgevende zaak is.

Het aantal studenten dat ten onrechte een onvoldoende krijgt hangt, zoals uit tabel drie af te lezen valt, nogal af (*ceteris paribus*) van het succes van maatregelen om te bewerken dat alleen studenten opkomen die in feite hun zaakjes voldoende beheersen. De percentages zakken drastisch naarmate het percentage dat in feite de stof voldoende beheerst, toeneemt van 75 naar 90 naar 95 (zie bijv. varianten 1, 2, en 3 waarin afwijzingspercentage en percentage in feite onvoldoende beheersing aan elkaar gelijk zijn). Het in de afweging betrekken van de groep ten onrechte onvoldoende beoordeelden wordt dan ook belangrijker naarmate door een fors afwijzingsbeleid meer studenten de stof moeten herhalen.

De mogelijkheden om de zakslaag grens op lagere percentages afwijzingen te laten uitkomen, hangen af van:

a) het succes van maatregelen om te voorkomen dat'. studenten lichtvaardig aan het tentamen deelnemen (terwijl zij in feite de stof niet voldoende beheersen); en

b) voorzover genoemde maatregelen degenen niet ontmoedigen deel te nemen die ondanks een behoorlijke studie inspanning de stof niet op tijd op voldoende niveau beheersten, hangt het af van het succes van beleidsmaatregelen die deze tekortkoming in het onderwijs kunnen wegnemen.

De eerste categorie is in het voorgaande op diverse plaatsen al uitgebreid aan de orde geweest, maar een toelichting op b) is hier wel op zijn plaats. Een en ander volgt natuurlijk uit uitgangspunt 4, zie blz. 14 e.v., en blz. 38 e.v.; aan de daar gegeven uiteenzetting valt het volgende toe te voegen:

Wanneer doelstellingen van onderwijs onderwerp van discussie zijn, wordt eigenlijk zonder uitzondering bij dergelijke discussies gerefereerd aan 'de normstudent', 'de gemiddelde student', de 'ruin

voldoende studentt, or what not. Dit is een uiterst primitieve wijze van denken.

Eén van de eerste dingen die de student in een cursus statistiek leert, is dat voor een enigszins zinvolle beschrijving van wat voor verschijnselen, dingen, of groepen personen ook, naast gegevens als gemiddelde o.i.d. ook gegevens over de spreiding nodig zijn. Het heeft weinig of geen zin om over de 'gemiddelde student' te spreken, zonder daarbij te betrekken de spreiding in studieresultaten zoals die kenmerkend is voor deze groep studenten in haar geheel, of over de spreiding in studieresultaten die wenselijk is, nog juist toelaatbaar is, e.d.

Afspreken dat onderwijsdoelstellingen refereren aan de 'gemiddelde student' is op zijn best een ambigue aangelegenheid.

Waar het om doelstellingen van onderwijs gaat, kunnen' we echter nog een heel stuk verder gaan dan de kritische aantekeningen die ik net gemaakt heb. Het is aan te tonen dat onderwijs dat niet op volstrekt absurde wijze is opgezet, er altijd in zal resulteren dat verschillen tussen studenten in termen van studieprestaties, studiemotivatie en attitudes e.d., door het onderwijs vergroot worden. Als het vergroten van dergelijke verschillen niet expliciet in de onderwijsdoelstellingen is opgenomen, zou in ieder geval bij de evaluatie van dat' onderwijs een aanwijzing voor het behoorlijk functioneren van dat onderwijs geput kunnen worden uit het controleren op het punt van die al dan niet groter geworden verschillen.

In ieder geval is het altijd een volkomen foute zaak om onderwijsdoelstellingen te formuleren voor een heel bepaalde doorsnede uit de studentenpopulatie, of voor een abstracte figuur als een 'gemiddelde' etc. student; In iedere groep studenten die het onderwijs beëindigt zullen de verschillen in studieprestaties groot zijn; de relatieve positie die een student in de rangorde op studieprestaties inneemt, is op zich ook nog weinig of geen indicatie voor de mate waarin onderwijs bij deze student in een stuk vermeerdering van kennis en inzichten heeft geresulteerd: dat hangt immers ook af van zijn voorkennis, capaciteiten, en motivatie aan het begin van dat stuk

onderwijs. Juist in de groep studenten die een soort 'gemiddelde prestatie' levert zul je figuren van diverse pluimage tegenkomen: studenten die gezien hun capaciteiten een lamelndige prestatie hebben geleverd, naast studenten die ondanks een gebrekkige vooropleiding o.i.d. toch met behoorlijke resultaten uit de bus zijn gekomen.

Aldus, verre van verschillen tussen studenten te verwaarlozen, zullen ze altijd expliciet betrokken moeten worden in de argumenten voor een beleidsbepaling inzake dat onderwijs.

Met name moet juist in onderwijs situaties waardering worden uitgesproken voor de student die met de stof meer moeite heeft dan anderen, en die er ondanks een heel behoorlijke studie-inspanning niet in slaagt om in de buurt van zoiets als de 'gemiddelde prestatie' te komen. Dat wil zeggen dat bij de cesuurbepaling ook de ruimte gelaten moet worden aan deze categorie studenten, voorzover althans zij rechtmatig aan het onderwijs deel mochten nemen, en zich de nodige inspanningen getroost hebben, zodat er geen enkele redelijke grond aanwezig is om hen een voldoende beoordeling te onthouden (behoudens een enkele onvermijdelijke foute beoordeling voortvloeiend uit de altijd onvolmaakte betrouwbaarheid van de toetsing en de beslissingsprocedure). De minimum eisen moeten dan ook zo geformuleerd zijn dat in principe (dat is: bij een behoorlijke studie-inspanning) alle studenten die tot dat onderwijs toegelaten zijn, binnen de tijd die daarvoor in het onderwijsprogramma is uitgetrokken, voor dat vak moeten kunnen slagen met' tenminste een voldoende beoordeling.

Voordat ik nog eens kort inga op de beleidsmaatregelen die overwogen kunnen worden wanneer kennelijk voor een belangrijke groep studenten geldt dat zij binnen die beschikbare tijd niet aan de gestelde eisen kunnen voldoen, een korte uitweiding over de relatie tussen wat boven gezegd werd over spreiding van studieprestaties en de idee van mastery learning of 'leren voor beheersing' van Bloom.

In regelrechte tegenstelling tot het boven door mij geponeerde lijkt de filosofie van Benjamin Bloom te staan (zie Bloom, Hastings, & Madaus 1971) dat onderwijzen nu juist moet betekenen alle studenten

tot het niveau van beheersing van de stof te brengen. Tot hetzelfde niveau dan ook. De geschriften van Bloom zijn hierover niet al te duidelijk, maar velen hebben er inderdaad uit gelezen dat dat dan ook in vrijwel alle onderwijs mogelijk zou moeten zijn, als je het maar goed aanpakt. Namelijk: dat je in dezelfde daarvoor ter beschikking staande tijd alle leerlingen of studenten, ongeacht verschillen in aanvangsniveau, tot eenzelfde eindniveau van beheersing van de stof kunt brengen. Ik geloof niet dat deze interpretatie kan steunen op enig canoniek geschrift van de heer Bloom, maar helaas is het een naleve opvatting die niet duidelijk door die canonieke schrifturen weersproken lijkt te worden. Het is wél zo, dat Bloom staande houdt dat verschillen tussen studenten aan het eind van de studie geen verband houden met verschillen tussen studenten aan het begin van de studie, en dat daarin mogelijkheden liggen om ongeacht verschillen in aanvangsniveau al die studenten tot eenzelfde eindniveau van beheersing van de stof te krijgen. Hij drijft dit principe natuurlijk niet tot in het absurde door, en laat daarbij bovendien ruimte voor verschillen in studietijd die studenten nodig hebben om datzelfde eindniveau te bereiken. Met het toelaten van verschillen in benodigde studietijd zijn we dan al een paar stappen verder gekomen: het is een beetje een behoorlijk kunstmatige zaak om studieprestaties op zich te onderscheiden van de tijd die nodig is geweest om het betreffende niveau te bereiken. Hoewel in het Nederlandse onderwijs dezelfde ingebakken gewoonte bestaat om veelal alleen te letten op de cijfers die behaald zijn, en niet op de (lange) tijd die de student nodig heeft gehad om dat voor elkaar te spelen. Ondanks dat alles blijft de visie van Bloom veel te optimistisch, voorzover het betreft de mogelijkheid om met 'goed' onderwijs bestaande verschillen tussen studenten tenminste niet te vergroten, en misschien zelfs te verkleinen. Behoorlijk bewijsmateriaal voor deze opvatting bestaat er niet. Dat wil zeggen, Bloom heeft begin van de zestiger jaren een opmerkelijke studie verricht waar hij concludeerde (Bloom, 1964) dat op vele gebieden verschillen tussen personen op het ene moment geen verband houden met verschillen tussen diezelfde personen zeg enige jaren later. Deze conclusie is echter gebaseerd op een twijfelachtige methodologie, die inderdaad ook herhaalde malen aangevochten is, maar waar jammer genoeg Bloom nooit expliciet op gereageerd heeft (Walberg 1972).

De gehele argumentatie hier komt er, evenals uitgangspunt 4, op neer dat de minimeisen zo geformuleerd worden dat de zwakkere studenten mits zij zich behoorlijk inspannen, binnen de voorgeschreven tijd een voldoende moeten halen. Er zijn drie argumenten waarom een dergelijke aanpak in ieder geval niet afgewezen hoeft te worden:

a) de problematiek van de zak-slaagbeslissingen is een geheel andere dan die van het differentiëel belonen van juist de 'betere' studenten;

b) cesuurbepaling heeft niet als doel om een stuk selectie onder studenten te plegen; d.w.z. voorzover een deel van deze 'zwakkere' studenten (maar zie puntje c.) inderdaad 'te zwak' zijn voor de studie, behoort dat niet op grond van studieprestaties voor dit vak alleen te worden uitgemaakt, en behoort dat zeker ook niet door een enkele docent of vakgroep uitgemaakt te worden (zie blz. 21 en volgende).

c) de 'zwakkere studenten' voor het ene vak zijn niet noodzakelijk ook de 'zwakkere studenten' voor andere vakken.

Dan nu een korte bespreking van mogelijke beleidsmaatregelen wanneer ontdekt of vermoed wordt dat een niet te verwaarlozen aantal studenten dermate grote moeite met de eisen heeft, dat ze niet binnen de beschikbare tijd met deze stof op voldoende niveau klaar kunnen komen. Al naar gelang de onderwijssituatie zullen bepaalde maatregelen meer voor de hand liggen dan andere. Een aantal mogelijkheden:

a) zonder meer verlagen van de minimum eisen.

b) de minimum eisen zijn vrijwel altijd op geheel verschillende manieren formuleerbaar. Om twee uitersten te noemen: de minimeis kan zijn dat de hele stof tenminste op bepaalde oppervlakkige wijze gekend moet worden (gelezen moet zijn bijv..) tegenover een formulering die bepaalt dat tenminste een met name genoemd deel van de stof zeer grondig beheerst moet worden. Er zou dus gezocht kunnen worden naar een 'Andere minimum formulering' waaraan ook de zwakkere studenten kunnen voldoen. (Zie ook blz. 31).

c) er is binnen het studieprogramma te weinig tijd voor dit vak uitgetrokken. In welk geval de docenten niet op eigen houtje de praktisch benodigde tijd gaan verhogen door de de eisen te verzwaren, maar zij in subfaculteitsoverleg kunnen proberen meer tijd beschikbaar te krijgen.

d) wanneer bij onderzoek blijkt dat het een bepaalde categorie studenten, met zeg een gebrekkige vooropleiding voorzover het dit vak betreft, is die in tijdnood komt, kan overwogen worden om voor deze studenten een ander programma op te zetten, afzonderlijke eisen te formuleren, of andere regelingen te treffen.

Het gaat er maar om, dat het aantal studenten dat in feite de stof onvoldoende beheerst op het moment dat het tentamen afgenomen wordt zo klein mogelijk gehouden wordt. Hoe kleiner dat aantal wordt, des te kleiner kan ook het aantal ten onrechte als onvoldoende beoordeelde studenten worden (zie tabel 3) voorzover het lukt procentueel ongeveer evenveel studenten af te wijzen als in feite de stof onvoldoende beheerst (maar met inachtneming van het op blz. 46 t/m 43 gestelde).

Als nu dat aantal studenten dat een onvoldoende krijgt inderdaad heel klein wordt, verliest de vraagstelling lover laten doen of niet' een groot deel van zijn praktische betekenis. Ook dan echter blijft het maar het beste om niet te eisen dat onvoldoendes ingehaald worden, maar de onvoldoende gewoon als onvoldoende te laten staan, en als zodanig ook in de examenregeling te betrekken.

Voorzover het niet lukt, om wat voor redenen dan ook, om het percentage onvoldoende beoordelingen klein te krijgen, is het de vraag nog steeds of het dan een juist beleid is om overdoen te verlangen op grond van de bewering dat een alsnog behaalde voldoende toch aantoont dat de stof dan ook beter beheerst wordt. Dat was de strekking van de stelling (argument 1) waar het hier nog steeds om gaat.

Hoe je het verder ook mag bekijken, we hebben bij beantwoording van de vraag te maken met drie verschillende groepjes studenten:

a) degenen die de eerste keer in feite de stof voldoende beheersten, maar op de toets een onvoldoende score boekten: voor deze studenten betekent de operatie 'overdoen' louter een verliespost. En dat geldt niet alleen voor hen persoonlijk, maar ook voor het onderwijs is het een verliespost deze categorie studenten nor een keer de zaak te laten bestuderen.

b) degenen die ondanks alles willens en wetens met onvoldoende voorbereiding opkomen voor de eerste tentamenafname, en daar een onvoldoende krijgen (verdiend dus). Als zij het de tweede keer al beter doen, en in feite zich dan ook beter geprepareerd hebben, is dat geen verdienste van het stelsel 'overdoen', maar is hier sprake van een tekortkoming in het onderwijs: deze groep studenten had immers de eerste keer helemaal niet op moeten komen.

c) degenen die binnen de beschikbare tijd de stof niet voldoende onder de knie konden krijgen. Het is natuurlijk mooi als zij de tweede keer de stof beter beheersen (zij hebben er dan misschien wel voldoende tijd aan kunnen besteden), maar dan is het allereerst de vraag of het onderwijs hier niet tekort geschoten is omdat ofwel de eisen te hoog gesteld waren, ofwel de beschikbare tijd te kort gehouden e.d. Dan nog zou overwogen moeten worden dat het stelsel van overdoen in het algemeen méér studietijd verlies met zich meebrengt dan bij meer soepele beoordelingsstelsels het geval hoeft te zijn, bijv. wanneer continue de gelegenheid bestaat tot het, afleggen van tentamens (niet alleen bij dit studieonderdeel, maar ook voor andere vakken in het programma) zodat de student tentamen aflegt op het moment dat hij de stof voldoende denkt te beheersen (o.a. mogelijk bij computer managed testing). Ten tweede, ondanks het voorgaande, is het niet direct erg waarschijnlijk dat de 'leerwinst' voor deze categorie studenten in de balans opweegt tegen de ellende die het systeem met zich meebrengt voor de studenten die ten onrechte een onvoldoende beoordeling hadden gekregen. Men moet al een bijzonder groot gewicht toekennen aan het stukje meerdere kennis van deze categorie studenten (tenslotte wisten zij bij de eerste tentamen afname ook al het een en ander van de materie af zij het dan ook niet voldoende), om de onnodige studiebelemmering en

tijdverlies voor de studenten die het ten onrechte nog eens overdoen als van geringer belang te kunnen beschouwen.

Het bovenstaande is wel voldoende argumentatie voor de juistheid van argument 1. Men moet daarbij wel bedenken dat de toelichting niet steunt op het uitgangspunt dat in principe iedereen een voldoende zou moeten behalen. Ook als men die opvatting niet meent te kunnen delen, blijft de gegeven argumentatie haar kracht behouden; alleen veranderen er hier en daar wat cijfers, zoals met behulp van tabel drie ingezien kan worden.

Toch is de toelichting niet compleet zonder er nog liet volgende aan te voegen:

(Aantekening bij punt e') hierboven: een belangrijke consequentie van de eventuele opvatting dat heb niet zo erg zou zijn om een student ten onrechte nog een keer het tentamen te laten doen omdat de leerwinst, voor degenen die het 'terecht' overdoen daar ruimschoots tegen op weegt, is dat volgens dezelfde redenering dan ook de zakslaag grens opgehoogd moet worden (zie blz. 4643) zodat meer studenten ten onrechte een onvoldoende gaan krijgen en wat meer studenten die de stof niet voldoende beheersten ook een terechte onvoldoende krijgen. Maar juist deze maatregel doet een stuk van het effect (van de leerwinst die men meent te verkrijgen in het stelsel van overdoen) teniet omdat het aantal ten onrechte onvoldoende beoordeelde studenten bij dit beleid veel sneller toeneemt dan het aantal 'terecht' als onvoldoende beoordeelde personen. Toch kan men er niet onderuit, omdat de ratio vereist dat men in zijn opvattingen en beleid coherent is: als men vindt dat overdoen ondanks alles de moeite waard is, dan moet volgens dezelfde argumentatie de aftestgrens verhoogd worden, waardoor juist een deel van, of de hele, winst verloren gaat. Tol'. zover deze aantekening.

Het hele voorgaande betoog is waarschijnlijk als onverwacht overgekomen, omdat de lezer op grond van de formulering van het argument 1 meer voorbereid was op een betoog dat een betere beheersing van de stof niet aantoonbaar zou zijn. Wat ik eigenlijk gedaan heb, dat is laten zien dat er een groep is voor wie geldt dat

zij de stof niet beter hoeft te beheersen dan de eerste keer (zij die ten onrechte een onvoldoende gekregen hadden) en een groep die de stof waarschijnlijk wél beter beheerst dan de eerste keer, maar dat

a) daaruit niet geconcludeerd mag worden dat laten overdoen dan ook een goed beleid zou zijn, omdat de oorzaken van het de eerste keer onvoldoende maken van het tentamen daarmee onaangeroerd blijven; en

b) die betere beheersing van de ene groep waarschijnlijk niet de moeite waard is, vergeleken met de onnodige ellende van de groep studenten die ten onrechte het tentamen overdoet.

Eigenlijk was dat dus iets anders dan in de stelling staat. Ik zal nu een andere toelichting op de stelling geven, maar ook deze toelichting 'past niet precies' op de stelling. Ik zal laten zien dat een verschil in feitelijke stofbeheersing moeilijk aan te tonen zal zijn, en dat ook in gevallen waarin dat eventueel wel zou lukken vervolgens nog maar aangetoond moet worden dat het verschil ook praktisch relevant is. Ik vind dat ik daarmee aan de stelling geen afbreuk doe, omdat het louter formeel kunnen weerleggen van de stelling (bijv. door een significant verschil in stofbeheersing aan te tonen) een Pyrrhus overwinning zou zijn wanneer dat verschil dan ook zo klein zou zijn dat er eigenlijk geen grond voor het beleid van 'overdoen' in gevonden kan worden.

Voor de groep die ten onrechte een onvoldoende kreeg, geldt dat zij bij hertoetsing onmiddellijk na het eerste tentamen een in veel gevallen ook statistisch aantoonbaar hogere score zullen behalen (hangt van de grootte van deze groep studenten af). Bij de normale herhaling van het tentamen zal voor deze groep gelden dat ook als zij de stof op hetzelfde, feitelijk voldoende, niveau beheersen als de eerste keer, zij een gemiddeld hogere score zullen behalen.

Merkwaardigerwijs doet hetzelfde effect zich ook bij de groep voor die terecht een onvoldoende kreeg, hoewel het effect hier veel kleiner is, mogelijk misschien in een experimentele situatie moeilijk aantoonbaar ook. Het hangt er een beetje van af hoe in een bepaalde situatie de cesuurbepaling in zijn werk is gegaan (hoeveel studenten

die de stof in feite onvoldoende beheersten toch een voldoende beoordeling kregen). Omdat dit effect heel klein is, zal ik het buiten beschouwing laten.

Bij elkaar bestaat de groep die het tentamen overdoet zeg voor ongeveer de helft uit personen die het de eerste keer in feite ook al voldoende kenden, en voor de andere helft uit personen die het de eerste keer wel voor een deel, maar in feite nog niet voldoende beheersten. Men mag natuurlijk wel aannemen dat voor deze tweede groep geldt dat zij de tweede keer hun zaakjes beter voorbereid hebben, en ipso facto zij de stof ook beter beheersen. In deze zin hoeven we eigenlijk geen bijzondere moeite te doen om aan te tonen dat er, in tegenspraak met stelling 1., wel degelijk een verschil in mate van beheersing van de stof zal bestaan. ik geef dan ook onmiddellijk toe dat het altijd mogelijk zal zijn, als het betreffende aantal studenten maar groot genoeg is, om aan te tonen, rekening houdend met het verschijnsel hierboven genoemd dat, zonder in feite de stof beter te beheersen voor een deel van de studenten geldt dat zij in ieder eval een hoger gemiddelde zullen halen dan de eerste keer (voor dat verschijnsel moet in de statistische analyse gecorrigeerd worden), dat voor de hele groep herhalers geldt dat zij gemiddeld de stof beter beheersen dan de eerste keer. Waar het op aan komt, is dat de mate waarin zij de stof aantoonbaar beter beheersen (de verhoging van het niveau dus), altijd klein zal zijn, bovendien zo klein dat men zich zal afvragen of dat verschil dan wel de hele trammelant van het laten overdoen waard is geweest.

(een conclusie uit dit verhaal is ook, dat op zich een gemiddeld hogere toetsscore nog geenszins ook hoeft te betekenen dat de feitelijke beheersing van de stof gemiddeld beter is geweest.)

Ik wil er nog aan toevoegen dat in het geval men er in zou slagen om een werkelijk relevant verschil in gemiddelde feitelijke beheersing van de stof tussen eerste en tweede tentamen aan te tonen, men daarmee ook een stukje onderwijsevaluatie heeft gepleegd dat er op neer komt dat ernstig gezocht moet worden naar de tekortkomingen in het onderwijs die dit grote verschil hebben mogelijk gemaakt. Want het is duidelijk dat in een dergelijk geval studenten de eerste keer

opkomen met zeer onvoldoende kennis van de stof, of dat studenten die eerst terecht of ten onrechte gezakt zijn, zich voor de tweede keer veel te goed op het tentamen voorbereiden (ten koste van tijd die aan andere studievakken besteed had kunnen worden), misschien omdat er een buiten proporties zware 'straf' staat op het twee keer gezakt zijn voor een tentamen. Naar andere achtergronden kan gezocht worden, al naar gelang er op basis van deze specifieke onderwijs inrichting vermoedens bestaan van bepaalde misstanden, misverstanden, of knooppunten.

De lezer mag wat mij betreft uit dit alles concluderen dat misschien stelling 1 in een aantal gevallen pertinent niet opgaat, maar dat precies in die gevallen gezocht moet worden naar verbeteringen in de onderwijssituatie die tot gevolg zullen hebben dat de stelling geldig wordt.

Argument 2. Voor studenten die ondanks behoorlijke voorbereiding een onvoldoende scoorden, zal een effect op verder studiesucces van het 'verbeterd' hebben van die onvoldoende, in gunstige zin niet aantoonbaar zijn.

Allereerst kunnen we het er snel over eens zijn dat voor de groep studenten die ten onrechte een onvoldoende beoordeling kreeg in ieder geval moet gelden dat het nogmaals doen van dit studieonderdeel alleen maar een ongunstig effect op verder studiesucces kan hebben. Voor wie dat niet onmiddellijk inzielt de volgende toelichting:

Wil je een effect, zoals hier bedoeld, onderzoeken, dan zou je daarvoor een experiment moeten uitvoeren ongeveer langs deze lijnen:

wil je onderzoeken of degenen die een onvoldoende prestatie ingehaald hebben door het vak over te doen, een betere studieprestatie leveren dan wanneer zij het niet overgedaan zouden hebben, dan is het duidelijk dat je deze groep random in tweeën moet verdelen, en dan de ene helft gewoon door moet laten gaan met de studie, en de andere helft het tentamen moet laten overdoen. (In de praktijk zullen nog een aantal kleine probleempjes moeten worden opgelost, omdat als het even kan de studenten zelf niet mogen weten

in welke groep (de experimentele of de controle) ze zitten, of wat de bedoeling van het experiment is, e.d. Maar daar hoeft bij deze uiteenzetting niet direct rekening gehouden te worden).

Vastgesteld moet worden wat onder 'studiesucces' verstaan gaat worden. Daarbij kan niet volstaan worden met gewoon maar zoiets als het gemiddelde behaalde cijfer te nemen. Allereerst niet omdat voor het behalen van eenzelfde cijfergemiddelde de ene student veel meer studietijd nodig gehad heeft dan de andere, zodat inderdaad met benodigde studietijd ook rekening gehouden moet worden. Anderzijds omdat in het huidige systeem de studenten zullen verschillen in de doeleinden die ze nastreven: de ene, student zal zo snel mogelijk willen afstuderen en daartoe mikken op telkens juist voldoende prestaties, de andere student wil met heel goede cijfers uit de bus komen en heeft daar een wat langere studietijd voor over, een derde categorie maakt erg veel werk van studieactiviteiten als het lopen van stages, geven van onderwijs, en schrijven van scripties en doen van onderzoek.

Ook al is het gelukt een in de meeste opzichten bevredigende 'criterium variabele' vast te stellen voor dat studiesucces waar het om gaat, dan is het vervolgens nog zaak om niet direct met deze variabele te gaan werken, maar met een variabele van aan de criterium variabele toegekende waarden (Zie Cronbach & Gleser 1965).

Eventueel ook zal men studiesucces willen 'corrigeren' voor gegevens als vooropleiding, verschillen in motivatie en capaciteit ten e.d., in de juiste veronderstelling dat geringe studieprestaties inderdaad weinig betekenen voor studenten van wie men meer had mogen verwachten, en veel kunnen betekenen voor studenten die qua vooropleiding en capaciteiten het met deze studie erg moeilijk hadden en er desondanks goed doorheen zijn gekomen.

Is dit alles goed overwogen, zijn meetinstrumenten en beoordelingsschalen ontwikkeld en de nodige gegevens verzameld, dan kan tenslotte nagegaan worden of de experimentele groep (zij die het tentamen overgedaan hebben) met hogere scores uit de bus komen dan zij die het niet overgedaan hebben.

Voordat besloten wordt een dergelijk omvangrijk onderzoek te gaan uitvoeren, moet de vraag beantwoord worden of dat onderzoek gegevens zou kunnen opleveren die iets toevoegen aan wat ook zonder het onderzoek uit te voeren al bekend is, en zo ja of die extra informatie consequenties kan hebben voor het te voeren onderwijsbeleid. Het gaat hier slechts om het principe dat praktisch onderzoek als dit ook beleidsgericht moet zijn, dat wil zeggen dat van te voren duidelijk moet zijn tussen welke beleidsalternatieven op basis van de te verkrijgen informatie een betere keuze gemaakt kan worden dan zonder het onderzoek nodig geweest zou zijn.

Ik zou er al deze woorden niet aan besteed hebben, wanneer ik er niet van overtuigd was dat in de meeste onderwijs situaties het uitvoeren van een dergelijk onderzoek, getoetst aan de genoemde criteria, een weinig zinvolle zaak is. Dat zal ik dan ook toelichten.

De alternatieve beleidsmaatregelen waartussen een dergelijk onderzoek zou moeten discrimineren zijn in dit geval het al dan niet handhaven van het instituut 'overdoen'. Ik zie er voor de onderhavige discussie van af dat er ook nog heel andere argumenten zijn om dat instituut af te schaffen, en beperk me tot argument 2) en het op blz. 71 geschetste soort onderzoek.

Omdat, benodigde studietijd niet buiten beschouwing gelaten kan worden, is het evident dat de groep studenten die het tentamen over moet doen omdat zij ten onrechte een onvoldoende beoordeling kregen, daardoor in hun studiesucces belemmerd worden. Voor deze deelgroep gaat argument 2. in ieder geval op; bovendien kunnen we opmerken, op grond van berekeningen zoals in tabel 3 dat het in de meeste gevallen zal gaan om een groep die ongeveer even groot is als de groep studenten die het tentamen 'terecht' overdoet; tenslotte, omdat voor de groep 'onterechte' herhalers een ongunstig effect op het studiesucces verwacht mag worden, moet het eventuele gunstige effect dat voor de groep 'terechte' herhalers gevonden wordt, daarvoor gecorrigeerd worden.

Het voorgaande is zonder verdere toelichting wel duidelijk genoeg.

Een veel moeilijker liggend punt is dat het onderzoek alleen maar verricht wordt voor degenen die ook inderdaad afstuderen, en dat al degenen buiten beschouwing blijven die de studie om wat voor reden ook vroegtijdig staken. Ook blijven buiten beschouwing de studenten die weliswaar een onvoldoende voor het tentamen behaalden, maar afstuderen zonder die onvoldoende overgedaan te hebben. De vraag die hier beantwoord moet worden is of door deze selectie van de betrokken groep studenten niet verwacht moet worden dat een eventueel zonder die selectie bestaand verband niet voor een aanzienlijk deel of helemaal verdwijnt. Het verschijnsel waar ik op doel is te illustreren aan de problematiek van eventuele toelatingsselectie voor het hoger onderwijs:

Tegen de tijd dat eventuele toelatingsselectie voor het hoger onderwijs aan de orde is, dat wil zeggen op het moment dat zich een aantal kandidaten aanmelden die allen met succes het voorgaande onderwijs tot en met het eindexamen hebben doorlopen, kandidaten die ook al een bepaalde studiekeuze hebben gedaan (daarbij rekening houdend met hun eigen belangstelling en capaciteiten), tegen die tijd mag je verwachten dat de eventuele voorspellende waarde van eindexamencijfers voor toekomstig studiesucces laag is of geheel ontbreekt. (Er zijn gevallen waarin zelfs negatieve verbanden gevonden zijn). Dat betekent niet dat eindexamen cijfers nietszeggende gegevens zijn; het betekent ook niet dat je geen betere voorspellingen zou kunnen doen wanneer het zo zou zijn dat alle leerlingen die eindexamen gedaan hebben universitair verder zouden studeren. Het betekent wel dat voor de groep die gaat studeren geldt dat verschillen in eindexamencijfers weinig of geen verband meer hebben met toekomstig studiesucces.

Hier is niets magisch' aan de hand. Het principe is dat studenten die eenmaal geselecteerd zijn op bepaalde gegevens, verder weinig of geen verband tussen diezelfde gegevens en de prestaties waarop de selectie gericht was, zullen laten zien. Een eenvoudige illustratie daarvan is het volgende waar gebeurde verhaal:

Voor de piloten opleiding in de Verenigde Staten tijdens wereldoorlog 2 meldden zich zeer veel kandidaten aan, waarvan echter

maar een klein deel geschikt was en geplaatst zou moeten worden. Het probleem was hoe uit die grote groep kandidaten de 'geschikten' te selecteren. Dat werd op de klassieke manier opgelost door iedereen een batterij tests af te nemen waarvan vermoed kon worden dat zij mogelijk succes in de opleiding zouden kunnen voorspellen. Om te weten te komen welke tests studiesucces voorspelden, werd de hele groep kandidaten tot de opleiding toegelaten, en werden de door hen behaalde studieprestaties vergeleken met de behaalde testcores. Uit die gegevens kon vervolgens een test batterij samengesteld worden die studiesucces bijzonder goed voorspelde, en op grond waarvan uit de grote groep nieuwe kandidaten voor de volgende opleidingsronde een relatief klein aantal 'geschikten' geselecteerd kon worden. Die aldus opgezette toelatingsprocedure functioneerde een tijd lang bevredigend. Totdat een nieuw aangetrokken psycholoog, die niet bij de ontwikkeling van de selectieprocedure betrokken was, eens ging onderzoeken wat nu eigenlijk het verband was tussen test scores en succes in de opleiding voor deze groep die juist op die tests geselecteerd was. En ja hoor, hij vond dat dat verband er vrijwel helemaal niet was, en concludeerde daaruit dat deze selectieprocedure maar onmiddellijk afgeschaft moet worden. Zijn conclusie was natuurlijk volstrekt onjuist: het ontbreken van voorspellende waarde binnen een op diezelfde tests geselecteerde groep zegt niets over de voorspellende waarde van de tests omdat die afgemeten moet worden aan de studieprestaties voor een groep toegelaten kandidaten die juist niet op diezelfde tests geselecteerd zijn!

Waarom is dat allemaal belangrijk voor onze onderhavige vraagstelling? Wel, het is niet onwaarschijnlijk dat de docent, op basis van zijn praktijkwaarnemingen en common sense redenties, in dezelfde val tuint. Die val is, dat voor studenten die afstuderen het niet onwaarschijnlijk is dat er geen verband is tussen studieprestaties in het begin, en aan het eind van de studie. In het bijzondere geval dat hier besproken wordt: het is niet onwaarschijnlijk dat er geen verband gevonden kan worden tussen het aldan niet overgedaan hebben van een 'terecht' onvoldoende beoordeeld studieonderdeel en later studiesucces voor degenen die de studie metterdaad afmaken.

Het laatste wil dus in het geheel nog niet zeggen dat slechte

tentamenprestaties geen voorspellende waarde hebben voor later studiesucces voor de hele groep studenten die die slechte prestaties boekte. Hoewel ook dat niet noodzakelijk zo hoeft te zijn.

Wanneer cesuurbepaling en selectie eindelijk van elkaar losgemaakt zijn, zou in bepaalde gevallen aangetoond kunnen worden dat bepaalde studieprestaties in het begin van de studie betekenen dat de kansen om de studie af te kunnen maken gering zijn, in welk geval de student aangeraden zou moeten worden de studie te staken. Maar juist de mate waarin studenten dan ook de studie staken, in die mate zal voor de overblijvende studenten gelden dat er een gering of geen verband tussen eerdere en latere studieprestaties is. Dit is eigenlijk alleen maar een extra pleidooi om, als men al wil selecteren en daar goede gronden voor heeft, die selectie uit te voeren los van de wijze waarop cesuren voor afzonderlijke vakken bepaald worden, en onbelemmerd door het instituut 'overdoen'.

(Zie Dawes 1975 voor een illustratie uit het Amerikaanse hoger onderwijs van dezelfde 'paradox').

Misschien ben ik nog niet duidelijk genoeg geweest. In andere woorden is wat ik bedoel het volgende:

Als het instituut 'overdoen' verdedigd wordt omdat voor studenten die 'terecht' een onvoldoende beoordeling kregen dat overdoen een gunstig effect op verder studiesucces zou hebben, dan zal men zich moeten realiseren dat deze uitspraak alleen maar zinvol zou kunnen zijn voor studenten die ook inderdaad afstuderen. Degenen die de studie toch niet aankunnen hoeven niet geplaagd te worden met opgelegde herhalingen, omdat die geen informatie toevoegen over hun capaciteiten. (Ten beste is het resultaat dat zij van het betreffende vak iets meer afweten). Voor de overigen geldt dat zij de studie mét (experimentele groep) of zonder (controle groep) overdoen af zouden maken. Binnen de laatste selecte groep is het niet bijzonder waarschijnlijk positieve effecten van dat overgedaan hebben te vinden, zeker niet waar het bovendien nog zo is dat de controle groep het niet te onderschatten voordeel heeft geen extra maanden studietijd kwijt te zijn aan het overdoen (studietijd die immers mee betrokken moet

worden in de maat die gehanteerd wordt voor 'studiesucces'). Voor wie ondanks dit alles toch nog meent een klein stukje 'winst' te moeten verwachten voor de 'overdoeners' moet dat nog afwegen tegen het 'verlies' voor diegenen die het tentamen ten onrechte hebben moeten overdoen en daardoor een handicap opgelegd kregen.

Een super simpele redenering die het paradoxale van de verwachting van hoger studiesucces door het laten overdoen illustreert, is wel de volgende:

In het op blz. 71 beschreven experiment zijn er een experimentele (over gedaan hebben) groep en een controle groep (niet overgedaan hebben ondanks onvoldoende beoordeling, een behandeling gegeven in het kader van dit experiment). Hoewel we niet in staat zijn de studenten die ten onrechte onvoldoende beoordeeld waren te onderscheiden van de studenten die terecht een onvoldoende kregen, weten we wel dat laten we zeggen voor ongeveer de helft van ieder van de groepen geldt dat ze ten onrechte een onvoldoende beoordeling hadden (zie tabel 3). In gedachten kunnen we dan ook vier groepen onderscheiden:

A studenten in de controle groep die ten onrechte een onvoldoende hadden gekregen

B studenten in de controle groep die terecht een onvoldoende hadden gekregen

C studenten in de experimentele groep die ten onrechte het tentamen hebben overgedaan

D studenten in de experimentele groep die 'terecht' het tentamen hebben overgedaan.

Wat mogen we redelijkerwijs verwachten met betrekking tot het gemiddelde studiesucces van deze groepen ? Zij het studiesucces van de bovengenoemde groepen gelijk aan A, B, C, en D respectievelijk. Dan is redelijk te veronderstellen dat het studiesucces van studenten die ten onrechte een onvoldoende kregen maar het niet overdeden iets hoger of minstens hetzelfde zal zijn dan dat van studenten die ten onrechte een onvoldoende kregen en het wél overdeden:

(i) $A \geq C$.

Zou (1) niet waar zijn, dan is het kennelijk een goed beleid om studenten altijd dit tentamen nog een keer te laten doen, een conclusie die in strijd is met iedere behoorlijke opvatting van onderwijsprogrammering.

Evenzo zijn de volgende relaties vanzelfsprekend:

(2) $A \geq B$.

(omdat je mag verwachten dat studenten die de stof in feite wel voldoende beheersten ceteris paribus in ieder geval geen slechtere verdere studieresultaten zullen behalen dan degenen die in feite de stof niet voldoende beheerste). Dus ook:

(3) $C \geq D$.

Uit (1) en (3) volgt dan ook nog:

(4) $A \geq D$. Of:

(5) $A \geq C \geq D$. (zie figuur 2 hier beneden)

Hoewel we weten dat $A \geq B$, is het nog een open vraag hoe B zich laat vergelijken met C en D. Een paar dingen kunnen we echter op voorhand wel zeggen: als A, C, en D dicht bij elkaar liggen, zal B daar ook heel dicht bij liggen; met andere woorden: als voor drie van de groepen verschillen in studiesucces nauwelijks bestaan, is er geen reden om te verwachten dat B een belangrijk stuk kleiner zal zijn dan D. Als het verschil $A - C$ erg groot is, betekent dat dat het overdoen van het tentamen voor de studenten die in feite de stof de eerste keer al voldoende beheersten, (studenten die alleen op grond daarvan een tenminste even goede studieprestatie verwachting hadden als studenten die in feite de stof niet voldoende beheersten) een bijzonder grote handicap opleverde. Naarmate die handicap groter is, is het onwaarschijnlijker te verwachten dat $B < D$ zal zijn, laat staan dat B tenminste evenveel kleiner zal zijn dan D als A groter is dan C (en dat

is wat je moet verwachten als mijn stelling onjuist zou zijn, en als beide betrokken groepen ongeveer even groot zijn). of $D \geq B \geq A \geq C$

Hoe het ook zij, een meer dan marginaal 'winstje' mag alleen in de allerbijzonderste omstandigheden van het 'overdoen' verwacht worden.

Dat B niet erg ver onder D kan liggen, als het er al onder ligt en niet er boven, wordt bevorderd doordat juist de studenten met verder in de studie de laagste prestaties, de studie niet zullen afmaken: dat forceert het gemiddelde voor de groepen in zowel de experimentele als de controle conditie omhoog, en wel méér naarmate de groep een lager gemiddelde heeft. Er is a.h.w. nauwelijks ruimte om B zoveel kleiner te laten zijn dan D als C kleiner is dan A. Dat is in onderstaande figuur ongeveer aangegeven door op de as van de studieprestatie ook het punt aan te geven dat als ondergrens functioneert: namelijk de minimaal vereiste studieprestatie Z om af te kunnen studeren. Z is een score waar de individuele student aan moet voldoen. Ieder van de door ons onderscheiden vier groepjes bestaat uitsluitend uit studenten die tenminste aan die norm voldaan hebben. Dat betekent dat de groeps-gemiddelden A, B, C en D tenminste een stukje boven Z moeten liggen, omdat voor de berekening van dat groeps-gemiddelde uitsluitend scores meedoen die groter of gelijk aan Z zijn. Het score bereik is met andere woorden éénzijdig afgekapt, en dat forceert de gemiddelden van deze vier groepjes dicht tegen elkaar aan omdat je wel mag aannemen dat alle vier groepjes niet erg veel goede en briljante studenten zullen bevatten vergeleken met die andere groep die voor het tentamen wel in één keer slaagde.

FIGUUR 2. Studieprestatieschaal. B ligt 'ergens' tussen minimum afstudeer eis Z (maar een stukje daarboven omdat B een gemiddelde is en A).

Tenslotte kan er nog aan toe gevoegd worden dat het hier niet gaat om vergelijking van de gemiddelde studieprestaties op zich, maar van de gemiddelden van de aan de individuele studieprestaties toegekende waarde of utiliteit, en daarvoor geldt nu juist in deze lage regionen

van de prestatieschaal dat het meestal heel weinig kan schelen of de prestatie wat hoger of wat lager is, zolang het gaat om prestaties die vlak bij het minimum toelaatbare niveau liggen. (Zie Cronbach & Gleser O.a. over toekenning van waarde aan criterium variabele scores). Ook dit werkt zo uit, dat A, B, C, en D op een kluitje gedrongen worden, vlakbij Z, zodat weerlegging van deze stelling 2 een bijzonder onwaarschijnlijke onderneming is.

Tot dusver heb ik nergens hoeven veronderstellen dat het overdoen op zich zou resulteren in een betere beheersing van de stof van het betreffende vak. Het is natuurlijk wel duidelijk dat de mate waarin ook dat niet het geval is, dat is de mate waarin men stelling 1 wil onderschrijven, het niet nodig is om stelling 2 nog eens afzonderlijk te beargumenteren. Hierboven heb ik echter laten zien dat een afzonderlijke argumentatie wel degelijk mogelijk is, ook al zou het voor de beheersing van de stof van het betreffende vak een 'goede zaak' zijn om het overdoen als gebruik te handhaven.

De gegeven argumentatie had betrekking op studiesucces in het algemeen, in tegenstelling tot bijvoorbeeld zoiets als prestaties voor een bepaald vak. Je zou namelijk ook kunnen redeneren, zoals vaak gedaan wordt, dat je dit vak toch wel tenminste op voldoende niveau moet beheersen om met vak zusendo in de verdere studie niet in de knoop te komen. De veronderstelling die dan gemaakt wordt is dat je voor een bepaald ander vak in de studie niet kunt slagen (of met veel moeite) wanneer je dit vak niet tenminste voldoende gemaakt hebt.

Hier hebben we weer zo'n voorbeeld van een heel populaire veronderstelling die altijd maar weer door docenten, studenten en andere direct bij het onderwijs betrokkenen gemaakt wordt: bepaalde stukken onderwijs moet je in een bepaalde volgorde doen, het ene vak is een voorwaarde voor het kunnen doen van het andere vak, of: als je een bepaald stuk noodzakelijke voorkennis mist is het altijd beter om die eerst maar eens bij te spijkeren en dan pas aan de reguliere cursus deel te nemen.

Bij onderzoek van de juistheid van dit soort opvattingen en veronderstellingen blijkt telkens weer dat er geen spat van deugt: de

veronderstelde effecten bestaan niet.

Voor wat dat bijspijker onderwijs betreft, ligt het misschien wel voor de hand als je de zaak zo bekijkt: het volgen dan dat bijspijkeronderwijs kost tijd, en het is misschien niet eens zo erg waarschijnlijk dat dat tijdverlies op gaat wegen tegen de betere prestaties die dankzij dat bijspijkeronderwijs misschien in de normale cursus behaald kunnen worden.

Voor cursus onderdelen die volgend de docent noodzakelijk in de ene en niet in de andere volgorde gegeven moeten worden, geldt waarschijnlijk in de meeste gevallen wel dat (tenzij je het helemaal te gek maakt, dus binnen de gewone onderwijs 'range' blijvend) het volgen van het meer geavanceerde stukje onderwijs vóór het in feite daarop voorbereidende onderwijs meer oplevert dan je naïevelijk zou verwachten omdat de student een enkel stukje extra benodigde informatie toch wel opscharrelt, geen onnodige tijd kwijt is met het tweemaal bestuderen van dezelfde onderwerpen (éénmaal op inleidend, en éénmaal op geavanceerd niveau) e.d.

Daarnaast heb je altijd nog die categorieën studenten aan wie het onderwijs in ieder geval niet besteed is, of je ze nu extra ondersteuning gegeven hebt of niet (degenen die de studie niet aankunnen, die omzwaaien, of om buiten de studie gelegen redenen er mee kappen).

Een uitstekende bespreking van de onderhavige verschijnselen is gegeven door Hills (1971). Een methode om de juistheid te onderzoeken van de veronderstelde noodzaak eerst bepaalde studieonderdelen gedaan te moeten hebben voordat je bepaalde andere studieonderdelen met succes zou kunnen doen, wordt gegeven door Pryor & Gordon (1974).

Met de laatste opvatting van 'overdoen is een voorwaarde voor verder studiesucces', en in mindere mate ook met de eerste opvatting dat 'overdoen verder studiesucces bevordert', is verwant de opvatting in andere regionen van ons onderwijs dat het 'zittenblijven' om de leerling voor nog groter calamiteiten te behoeden, om zijn kennis eerst op het niveau te brengen dat men voor volgende klassen noodzakelijk vindt, e.d. Ook het zittenblijven is een probleem waarmee we niet

alleen in Nederland worstelen, maar dat in vele landen problematisch is, en onderwerp van onderzoek is geweest.. Uit resultaten van onderzoek op dit gebied kunnen met de nodige voorzichtigheid ook conclusies getrokken worden ten aanzien van de mogelijke juistheid van veronderstellingen die ten grondslag kunnen liggen aan het instituut 'overdoen'. Een recent overzicht van de onderzoekliteratuur werd gegeven door Jackson (1975). In zijn inleiding maakt Jackson al duidelijk dat er geen rationele grondslag voor het doubleerbeleid lijkt te bestaan:

"It is not at all clear from good common sense whether either of these justifications accurately reflects what happens to a child when he or she is retained in a grade. There is little doubt that a pupil who is having serious academic difficulties in one grade is likely to continue having difficulties if merely promoted to the next grade. What is not clear is how the mere repetition of a grade of schooling is likely to reduce these difficulties. Very seldom is there any substantial special help provided to repeating pupils; instead, they are recycled through a program that was inappropriate for them the first time and that may be equally the second time inappropriate and of less interest to them.

In plaats van domweg laten herhalen, zou tenminste overwogen kunnen worden of meer op de problemen van deze studenten afgestemde onderwijsmaatregelen niet veeleer genomen zouden moeten worden. Een interessante beschouwing over verschillende ins and outs daarvan wordt gegeven door Salomon 1974. Waar het gaat om een vergelijking tussen de beide alternatieven overdoen of niet (doubleren of niet) geven de onderzoeksresultaten Jackson aanleiding tot de volgende concluderende opmerkingen:

"One general conclusion about the effects of grade retention relative to grade promotion is clearly warranted by all the results taken as a whole: There is no reliable body of evidence to indicate that grade retention is more beneficial than grade promotion for students with serious academic or adjustment difficulties."

"Thus, those educators who retain pupils in a grade do so without

valid research evidence to indicate that such treatment will provide greater benefits to students with academic or adjustment difficulties than will promotion to the next grade."

Jackson legt wat de onderzoek kant betreft, de nadruk op methodologische problemen, en waarschuwt tegen de vaak moeilijk te onderdrukken neiging om 'sloppy research' te doen, geen behoorlijke experimentele opzet (zie bijv. blz. 71) te hanteren: "Those investigators who are inclined to use a nonexperimental approach to analyzing the relative effects of grade retention and grade promotion should proceed very cautiously, if at all."

Er zijn nog een aantal opmerkingen te maken over de veronderstelling dat het behalen van een voldoende een minimale voorwaarde zou zijn voor het behoorlijk kunnen doen van één of meer vakken in het vervolgonderwijs. In het voorgaande is al uitgebreid geïllustreerd hoe onwaarschijnlijk het is om ooit een dergelijk verband aan te tonen; maar dan zou je vervolgens nog kunnen vragen welke conclusies er vallen te trekken voor het geval het verband wél aangetoond zou kunnen worden in een bepaalde onderwijs situatie. De conclusies liggen al min of meer voor de hand: naarmate het gaat om een omvangrijker groep studenten, kun je eigenlijk alleen maar concluderen dat voor het onderhavige vak in het studieprogramma dermate weinig tijd (te weinig) tijd is uitgetrokken dat voor een belangrijk aantal studenten helt, niet mogelijk is om binnen die tijd tenminste op zó'n niveau de stof te beheersen dat de verdere studievoortgang niet gehinderd wordt door een gebrekkige beheersing van het onderhavige vak.

Zo de lezer daar al eerder naar gezocht mocht hebben, ligt in het zojuist beschrevene ook de verklaring voor het in veel onderwijs situaties niet aan kunnen tonen van genoemde veronderstelde verbanden: als het onderwijs behoorlijk is opgezet, zullen die verbanden niet voorkomen. Een ander soort verklaring is te vinden in de vergaande plooibaarheid van de activiteiten van de student: hij is vaak bijzonder goed in staat om eventueel extra beslag dat door een voor hem 'moeilijk' vak op zijn tijd gelegd wordt, op te vangen door de benodigde tijd weg te halen bij

andere vakken die hij beter aankan, of bij vrijetijdsactiviteiten die hij tijdelijk kan opschorten. Zolang het onderwijsprogramma ruimte laat voor dergelijke 'opschuiverij', en ook door andersoortige keuzemomenten zoals de keuze uit verschillende onderwijsmethoden (zelfstudie, groepswork) en toetsingsvormen (schriftelijk, mondeling, computerized testing) inbouwt, worden ongelukkige voorwaardelijke koppelingen tussen verschillende studieonderdelen voor zover mogelijk vermeden.

Natuurlijk zijn er nog vele andere mogelijkheden om het 'gat' tussen populaire opvattingen over het functioneren van het onderwijs en de bij empirische toetsing daarvan nogal radicaal afwijkende werkelijkheid, te vullen. Een laatste nog te noemen ingang is dat men bij psychologische en sociale verschijnselen sterk geneigd is te generaliseren op basis van weinig en bovendien nog weinig betrouwbare informatie. Het is in het algemeen al een menselijke eigenschap om de eigen kennis te overschatten, ook waar het gaat om heel precieze informatie. Vraag bijv. eens een aantal personen om voor bepaalde landen het aantal inwoners te schatten op de volgende wijze: geef aan de twee getallen waarvan je denkt dat de kans $1/3$ is dat het werkelijke aantal er tussen ligt, en $1/3$ dat het beneden het kleinste aantal, resp. boven het grootste aantal ligt. Dergelijk soort experiment leidt tot resultaten waarbij blijkt dat niet in $1/3$ van de gevallen de werkelijke getallen tussen de aangegeven grenzen liggen, maar in ongeveer slechts $1/6$ van het aantal gevallen: een zeer forse overschatting van de eigen nauwkeurigheid dus. Welnu, ook in het onderwijs zijn de betrokkenen bij dat onderwijs voortdurend bezig om allerlei informatie over dat onderwijs te verzamelen, en dat gebeurt door schattingen te maken op basis van onnauwkeurige, en niet op validiteit gecontroleerde waarnemingen. Het is niet te veel gevraagd om de veronderstellen dat de nauwkeurigheid van opvattingen over onderwijs situaties zoals die bij docenten leven, zeer veel geringer is dan de nauwkeurigheid (ook al zeer gebrekkig) waarmee we menen veel duidelijker zaken te 'kennen'. Met deze wat naar het filosofische neigende opmerkingen besluit ik deze toch al vrij lang geworden toelichting op stelling 2.

Argument 3. Het laten overdoen van onvoldoende gemaakte

studieonderdelen veronderstelt impliciet dat het laten overdoen een betere procedure is dan het gewoon door laten gaan met de studie. Die aanname is in vrijwel alle onderwijs situaties aantoonbaar onjuist.

In de toelichting op argument 2. kwam af en toe al ter sprake dat het laten overdoen van studieonderdelen tenminste veronderstelt dat overdoen beter is dan andere alternatieven, zoals gewoon doorgaan met de studie, een ander onderdeel van hetzelfde vak bestuderen, hetzelfde nog eens bestuderen maar dan onder intensieve begeleiding, etc. Voor het laten overdoen van vakken geldt hetzelfde als wat Jackson (1975) zei over het doubleren:

"Thus, those educators who retain pupils in a grade do so without valid research evidence to indicate that such treatment will provide greater benefits to students with academic or adjustment difficulties than will promotion to the next grade."

In deze uitspraak ligt ook het oordeel besloten dat je in dergelijke omstandigheden af moet zien van het laten doubleren of overdoen, omdat er grote nadelen aan verbonden zijn voor de betrokken leerlingen of studenten, evenals voor het onderwijs (en de kosten van dat onderwijs). Die nadelen zijn wel voldoende bekend: herhalingen, en de voortdurende dreiging van het moeten herhalen als je geen voldoende zou krijgen, zetten de student onder druk, bedreigen hem in zijn persoonlijke en materiële zekerheden, etcetera, ook afgezien van eventuele voordelen op het punt van 'betere beheersing van het vak'.

Bij de verdere toelichting van dit argument zullen we er veel gemak van hebben wanneer gewoon over de 'opbrengst' van verschillende alternatieven voor het onderwijs beleid kunnen spreken. In dit geval zou dat zijn dat we de 'opbrengst' van het stelsel 'overdoen' vergelijken met de 'opbrengst' van het alternatief dat er niets overgedaan wordt, of dat er iets anders extra gedaan wordt.

Opbrengst kan in verschillende situaties en door verschillende personen en instanties natuurlijk anders ingevuld worden. Als inleidend voorbeeld kies ik bespreking van de situatie zoals de docent die zou kunnen zien, terwijl voor hem uitsluitend de 'opbrengst' in

termen van aantal beheerste doelstellingen van belang is.

Veronderstel dat de stof voor een bepaald studieonderdeel zich op laat delen in een relatief groot aantal onderdelen, waarbij in de toetsing op tamelijk duidelijke wijze te schatten is hoeveel onderdelen door de student 'beheerst' worden. Deze aanname laat onverlet dat er bij de zakslaag beslissingen nogal wat fouten gemaakt worden, zoals in de bekende tabel 3 aangegeven. Veronderstel dan dat voor de docent de rechtvaardiging voor het laten overdoen er in gelegen is dat deze studenten bij het in tweede instantie slagen tenminste een minimum aantal onderdelen uit de stof beheersen. Het overdoen kost deze studenten even veel tijd als ze anders kwijt geweest zouden zijn aan het doen van een ander vak uit het studieprogramma. Gezien deze toch wel ernstige consequentie van zijn zakslaag beleid heeft de docent besloten om bij dat beleid te overwegen in plaats van de studenten zijn eigen vak te laten overdoen bij een 'onvoldoende' beoordeling op het eerste tentamen, hen een extra keuzevak te laten doen, wanneer aannemelijk gemaakt kan worden dat het laten doen van dat extra keuzevak er in resulteert dat de studenten per saldo meer onderwerpen beheersen, d.i. meer doelstellingen bereikt hebben.

Deze docent realiseert zich ook, dat hij bij het plegen van die analyse onderscheid moet maken tussen een aantal deelgroepen:

1. degenen die ten onrechte een onvoldoende kregen;
2. degenen die in de beschikbare tijd zich onvoldoende konden voorbereiden, en ook een onvoldoende kregen;
3. degenen die om andere redenen onvoldoende voorbereid opkwamen, en een onvoldoende kregen.

Voor categorie 3. geldt dat het niet erg zinvol is om het al dan niet laten overdoen van het tentamen als alternatieven te bespreken, omdat hier het duidelijkste alternatief is het onderwijsbeleid zo te wijzigen dat deze studenten ontmoedigd worden om met onvoldoende voorbereiding aan tentamens deel te nemen. Maatregelen om dat te bereiken zijn tamelijk eenvoudig en al eerder

aan de orde geweest (o.a. verhogen van de transparantie van de toetsingssituatie, hanteren van een soort 'grade point average' systeem, frequenter gelegenheid geven het eind tentamen af te leggen, e.d.); afgezien van de voordelen die dergelijke maatregelen op zich al inhouden, is een belangrijk resultaat van het slagen van een dergelijke strategie dat het nemen van zak-slaagbeslissingen nu met minder fouten kan gebeuren (zie tabel 3 voor wat de gevolgen zijn wanneer het aantal studenten dat de stof in feite onvoldoende beheerst afneemt).

Daarmee zijn we nog niet klaar met deze categorie 3. Immers, het gebruik om onvoldoende gemaakte tentamens over te laten doen kan op zich al met zich mee brengen dat studenten aangemoedigd worden om een gokje te wagen wanneer ze in feite zeer onvoldoende op het tentamen voorbereid zijn. Een alternatief stelsel waarin geen herhalingen meer gevraagd worden, zal het wagen van dit soort gokjes ontmoedigen omdat de dan behaalde 'onvoldoendes' ook al 'onvoldoendes' zullen blijven staan in het dossier van de student. Met andere woorden: bij het plegen van de analyse kan groep 3 weliswaar buiten beschouwing blijven, maar dan met de aantekening dat een ander beleid (of dat nu gewoon door laten gaan met de studie is, of het laten doen van een ander, extra, studie onderdeel maakt hier niet uit) voorzover het deze groep betreft, gunstiger zal zijn.

Voor degenen die ten onrechte een onvoldoende kregen, geldt dat zij er niets mee opschieten hetzelfde vak nog eens over te doen. Misschien zullen ze de stof de tweede keer beter beheersen dan de eerste keer, omdat ze minder risico willen lopen weer een onvoldoende te krijgen, maar ook dan blijft het feit dat zij de eerste keer de stof al voldoende beheersten, en dat is alles wat gevraagd wordt. Het is mooi, wanneer studenten méér doen dan de minimum vereisten, maar voor het zakslaag beleid gaat het alleen en uitsluitend om die minimumvereisten. Het is dan ook onzuiver om te redeneren dat de studenten die ten onrechte een vak zitten over te doen, daar in veel gevallen dan toch wel wat extra van op zullen steken. Dezelfde redenering geldt voor alle studenten, en het zou absurd zijn om alle studenten de zaakjes twee keer te laten doen. Een dergelijk argument gaat iedere doelstelling van onderwijs ver te buiten.

Daarom zullen we voor deelgroep 1. aanhouden dat bij overdoen zij geen enkel onderwerp meer zullen beheersen dan zij al deden bij de eerste tentamen afname. Voor wat het alternatief betreft, kunnen we wel aannemen dat zij gemiddeld een 'normale' prestatie zullen leveren en het 'normale' aantal onderwerpen op het afsluitende tentamen voor dat vak zullen beheersen.

Een en ander is uitgetekend in figuur 3 op de volgende bladzijde.

FIGUUR 3. Opbrengst van overdoen vs. extra vak voor ten onrechte ais onvoldoende, resp. terecht als voldoende beoordeelden.

[noot 2002: twee horizontaal verlopende functies, extra vak boven overdoen op verticale opbrengst-schaal.]

De figuur is zo simpel dat het misschien niet helemaal meer voor de hand ligt welke conclusies daaruit te trekken zijn. De opbrengst var. het doen van een 'extra' vak is over de hele linie hoger dan de opbrengst van het over laten doen. In ieder geval voor deze deelgroep zou dan ook het beste beleid zijn om ze niet het betreffende vak over te laten doen, maar een ander vak als extra vereiste te laten doen. Andere alternatieven zijn natuurlijk ook nog te overwegen, zoals het gewoon door laten gaan met de studie, maar daarover straks meer.

De fiuur moet voor een goed begrip nog verder toegelicht worden. Op de horizontale as staat de score die behaald is op het eerste tentamen. De figuur is dan als volgt te lezen: voor iedere mogelijke behaalde tentamenscore lezen we verticaal af wat voor een student met die tentamenscore verwacht mag worden aan 'opbrengst' wanneer hij het tentamen overdoet, dan wel wanneer hij een extra vak kiest als 'compensatie' voor deze onvoldoende.

(De figuur kan ook zo gelezen worden: veronderstel dat iedere deelgroep die een bepaalde score op het tentamen behaalt, random in tweeën gedeeld wordt, en de ene helft het tentamen gaat overdoen, de andere helft een extra vak doet. De figuur geeft dan de verwachte opbrengst voor ieder van beide groepen. En dat voor ieder van de behaalde tentamenscores.)

in de figuur is ook de verwachte opbrengst ingetekend voor degenen die op het eerste tentamen een voldoende beoordeling kregen. Ook voor hen kan namelijk nagegaan worden of de genomen beslissing wel juist is, en in ieder geval kan direct een vergelijking gemaakt worden tussen degenen die een onvoldoende, en degenen die een voldoende beoordeling kregen. Het lijkt een beetje absurd om het zo te doen, maar over dat absurde karakter van dit gekozen voorbeeld wil ik het nu juist hebben.

Voor degenen die ten onrechte een onvoldoende kregen, geldt dat zij tenminste beter een ander, extra, vak kunnen doen, dan dezelfde hap nog een keer te herkauwen. Dat 'extra' vak heb ik echter alleen maar ingevoerd om de docent zo ver te krijgen, dat hij erkent dat het in ieder geval voor deze studenten beter zou zijn dat extra vak te doen, dan het vak over te doen. Maar diezelfde conclusie zou ook gelden voor degenen die terecht een voldoende kregen! En de laatste groep laten we in ieder geval geen extra vak doen, want dat zij neerkomen op een reguliere uitbreiding van het studieprogramma; zij gaan met het gewone studieprogramma verder. Welnu, dan zou deze docent nu moeten concluderen dat ook ten onrechte onvoldoende beoordeelden beter gewoon door kunnen gaan.

FIGUUR 4. Opbrengst van overdoen vs. extra vak voor terecht als onvoldoende, resp. ten onrechte als voldoende beoordeelden.

[noot 2002 extra vak: oplopende curve; overdoen: dalende curve; maar zo dat beide elkaar niet snijden.]

Dan is bovenstaande figuur 4. voor deelgroep 2 als volgt toe te lichten:

Het gaat hier om studenten die in de programmatisch ter beschikking staande tijd zich onvoldoende voor dit vak konden voorbereiden. Als zij het tentamen overdoen, kunnen we ervan uit gaan dat zij zoveel onderwerpen er bij leren beheersen, als voldoende

is om een voldoende te behalen. Als zij meer zouden doen dan dat, zullen we dat ook hier niet op het credit van de overdoen vereiste schrijven, omdat het er hier slechts om gaat of studenten de stof op een bepaald minimum niveau beheersen of niet. De opbrengst van het over laten doen ligt dan ook over de hele horizontale as boven de nullijn, en we mogen er wel van uit gaan dat degenen met de lagere tentamenscores er meer bij hebben moeten leren dan de anderen, zodat de lijn schuin naar rechts afloopt. Door de bank genomen zullen deze studenten bij de eerste tentamen gelegenheid niet al te ver beneden het vereiste minimum niveau gezeten hebben (omdat zij alle voorgeschreven tijd besteed hebben aan de voorbereiding op het tentamen), zodat de mééropbrengst bij de herhaling niet al te groot kan zijn. Die verwachte opbrengst kan vergeleken worden met de verwachte opbrengst voor een geheel nieuw vak, en dat is precies de lijn 'extra vak' in figuur 4. Over deze lijn, die hier anders loopt dan dezelfde lijn in figuur 3, valt direct op te merken dat hij naar rechts oploopt, en net niet aan de overigens voor dat extra vak gewonegemiddelde opbrengst 'M1 toekomt. De reden daarvoor is de voor de hand liggende veronderstelling dat figuur 4. betrekking heeft op een groep studenten die waarschijnlijk niet alleen maar voor dit vak in studiemoeilijkheden komt, maar ook voor andere vakken iets meer moeite zal hebben met de stof dan de andere studenten: vandaar de verwachting dat voor deze groep de verwachte score iets lager ligt dan het overall gemiddelde, en wel des te meer naarmate de score op het tentamen lager was. De veronderstelling is dus dat er een correlatie bestaat tussen beide tentamens, zij het dat dat geen forse correlatie hoeft te zijn.

Desondanks loopt de opbrengst lijn voor 'extra vak' consistent overal boven de opbrengst lijn voor 'overdoen', wat zou moeten betekenen dat het ook voor deze groep studenten beter is ze een extra vak te laten doen i.p.v. te laten herhalen. Ofwel, omdat we dat extra vak tenslotte wel aan iedereen zouden kunnen vragen, het is maar het beste om te signaleren dat dit vak onvoldoende was, en ze gewoon door te laten gaan.

De conclusie zou, althans wat dit voorbeeld betreft, pas anders zijn wanneer beide lijnen elkaar ergens binnen het tentamenscorebereik

zouden kruisen. Dan zou het voor de studenten met een tentamen score links van dat punt beter zijn om het over te doen, en voor de anderen om gewoon door te gaan met de studie. De vraag is dan ook, of je kunt verwachten, in dit voorbeeld, dat de beide lijnen elkaar ooit zullen kruisen. Het antwoord daarop is Nee, en wel om de volgende eenvoudige reden:

Er is op voorhand geen enkele reden om te veronderstellen dat een student die om te beginnen al moeilijkheden met dit bepaalde vak had, er in zou slagen om bij een gewone herhaling van hetzelfde vak er meer bij te leren dan hij er bij had kunnen leren wanneer hij gewoon een ander vak in die tijd had bestudeerd. (Gewoon doorgegaan was met de studie bijvoorbeeld). Je zou je natuurlijk kunnen voorstellen dat dat andere vak veel moeilijker is, zodat deze studenten daar minder op zouden steken in termen van aantal beheerste onderwerpen, dan wanneer ze het eerste vak nog eens zouden herhalen. Het is echter niet redelijk om de veronderstelde 'moeilijkheid' van een vak niet in de beschouwing te betrekken, en in de opbrengst mee te wegen. Het resultaat van dat meewegen is dan weer dat beide lijnen elkaar niet zullen kruisen.

Het is nu langzamerhand tijd om van de beschouwing van dit voorbeeld over te gaan op de bespreking van het meer algemene geval. En dat is dat het in het algemeen zo is, dat wanneer studenten op grond van een toets score een verschillende behandeling krijgen, de veronderstelling gemaakt wordt dat de opbrengst (wanneer, zeg, iedereen ongeacht tentamenscore dezelfde behandeling krijgt. Zolang opbrengst lijnen binnen het relevante score bereik elkaar niet kruisen is het het beste om iedereen dezelfde behandeling te geven, namelijk die behoort bij de hoogst liggende opbrengst lijn. Pas wanneer beide lijnen elkaar binnen het scorebereik kruisen, is het inderdaad beter om de beide groepen anders te behandelen, en wel de groep links van het snijpunt volgens de behandeling die daar de hoogste opbrengst geeft, en de andere groep idem dito.

De veronderstelling die gemaakt wordt bij een onderwijsbeleid waarin studenten bepaalde vakken moeten overdoen, is dat de resultaten daardoor beter worden, dat de payoff onder dit beleid

groter is dan onder alternatieven, zoals die waarbij iedereen dezelfde behandeling krijgt. Een punt dat als één van de eersten gemaakt werd door Cronbach & Gleser (1965, blz. 251)

"Even more to the point is the fact that use of tests for educational placement implicitly assumes the existence of such different payoff functions. Our model therefore is merely a formal statement of what is everyday accepted without question, and examining its implications for testing is undoubtedly important. The realization that this assumption underlies practice is in itself an advance, because it makes clear the need for research on payoff functions* (opbrengst lijnen).

But let us first defend the claim that this assumption is currently made. Dividing a group on the basis of some univariate information and assigning each segment to a different treatment must be justified by the belief that greater payoff will be obtained in this manner."

FIGUUR 5. De impliciete veronderstelling m.b.t. de opbrengst lijnen die aan het laten overdoen ten grondslag ligt.

[noot; payoff curven snijden elkaar, op dat punt ligt de cesuur]

With two treatments, there are three possibilities:

1. The payoff functions are identical, in which case placement is valueless since the treatment to which a man is assigned makes no difference in payoff.

2. The payoff function for one treatment is uniformly higher than the other for all score levels. In this event it is unwise to divide the group. All persons should be given the superior treatment unless institutional constraints require both treatments to be used (e.g., where there are too few therapists to give psychotherapy to all patients).

3. The payoff functions intersect somewhere within the score range. If this is the case, (als in de figuur), it is profitable to assign men to different treatments. so that each receives the treatment that 'gets the most out of him'.

In de normale onderwijspraktijk zijn verbanden, zoals in de figuur getekend, natuurlijk niet bekend. Men is zich zelfs in het geheel niet er van bewust hoe de onderwijs payoff geëvalueerd zou kunnen worden. Het ontbreken van dergelijke gegevens belet ons overigens in het geheel niet om met recht te kunnen stellen dat het beleid waarin overdoen van studieonderdelen kennelijk niet ongewenst is, berust op de aanname van een verband zoals in de figuur ongeveer geschetst. En hoewel zoals gezegd de precieze empirische gegevens niet beschikbaar zijn, kunnen we een heel eind komen door eens te onderzoeken hoe waarschijnlijk het is dat in een normale onderwijssituatie aan deze vooronderstelling kan worden voldaan.

Het zal niet verbazen, dat een dergelijke analyse voor verschillende belanghebbende personen of groeperingen en instellingen afzonderlijk gedaan moet worden. Want het zal nog al enig verschil maken, of de docent de 'gene is die verschillende 'opbrengsten' met elkaar gaat vergelijken, of dat het de student is die dat doet, en dan hebben we daarnaast nog de subfaculteit die weer iets andere belangen heeft, en verantwoordelijkheden, en de overheid die eveneens regulerend kan optreden waar het de inrichting van onderwijs betreft, en de verschaffing van de nodige materiële middelen.

Als opbrengst of payoff op eenvoudige wijze in guldens uitgedrukt kan worden, is het een kleine moeite om na te gaan bij welk onderwijsbeleid de opbrengst het gunstigst is. Wanneer het niet direct mogelijk of zelfs maar zinvol is om uitkomsten van onderwijs in guldens uit te drukken, wil dat nog niet zeggen dat beleidsalternatieven niet op dezelfde wijze geëvalueerd kunnen worden als dat het geval zou zijn wanneer het om direct 'verhandelbare' producten zou gaan. Het wordt alleen maar iets lastiger.

Ook al is het voor onderwijs zaken niet zinvol om uitkomsten in guldens uit te drukken, en ook al wordt het afwegingsproces voor de beschikbare beleidsalternatieven daardoor bemoeilijkt, dan is het toch nog zo dat we daar pas echt last van krijgen op het moment dat het gaat om beleidsuitkomsten die zo dicht bij elkaar in de buurt liggen dat een nauwkeurige vaststelling van de opbrengst van de alternatieven

nodig is om een keuze te kunnen maken. In veel gevallen is het echter ruimschoots voldoende om op grove wijze de uitkomsten te evalueren, omdat een overzicht van het ene alternatief op het andere al snel duidelijk wordt.

Het laatste is ook het geval voor de afweging van het over laten doen van studieonderdelen, versus het gewoon door laten gaan met de studie. Bij de verdere analyse zullen we het dan ook relatief gemakkelijk hebben.

Een opmerking die voor de volledigheid hier wel gemaakt moet worden, maar die voor goed begrip van het volgende niet direct noodzakelijk is: er is een wezenlijk verschil tussen de directe resultaten van een bepaald beleid (gemeten in guldens, studieresultaten, bestede tijd, wenselijkheid van het bereikte, o.i.d., en het nut of de utiliteit van die opbrengst. Het nut van een extra verdiende gulden is er bijv. van afhankelijk of je al twee gulden verdiend had, of dat het er duizend waren; in het laatste geval kan die extra gulden ons "niets schelen, terwijl in het eerste die extra gulden een belangrijk resultaat kan zijn. Voor een zorgvuldige analyse is het dan ook nodig direct gemeten opbrengsten te onderscheiden van de waarde aan die opbrengsten toegekend. Voorzover mogelijk zal ik in het volgende het onderscheid niet expliciet maken, en gewoon spreken van 'opbrengst'.

De opbrengst van een bepaald onderwijsbeleid kan op verschillende manieren gemeten worden, waarbij het vaak niet mogelijk is om één van de mogelijkheden als de meest geschikte aan te wijzen, maar er met een combinatie van meerdere gegevens gewerkt moet worden. Waar het om onderwijs gaat, kun je direct denken aan gegevens als : bestede studietijd, behaalde studieresultaten, gerealiseerd hebben van de onderwijsdoelstellingen, kosten van het onderwijs in termen van docenten tijd, kosten van materiële voorzieningen, kosten van administratieve ondersteuning, e.d., vaak ieder op verschillende manieren vast te stellen. (Verschillende manieren: studieresultaten kun je op heel uiteenlopende manieren toetsen, met verschillende soorten van vraagstellingen; het bereikt zijn van diverse doelstellingen kan meestal slechts geëvalueerd worden door resultaten verkregen met verschillende onderzoekjes met elkaar te combineren; e.d.)

Er bestaat, waar het om evaluatie van onderwijs gaat, een maar moeilijk te onderdrukken neiging om daarvoor te grijpen naar gegevens die het makkelijkst te verkrijgen zijn, en een grote mate van 'face validity; oppervlakkige geldigheid, hebben: de cijfers die studenten behaald hebben. In het algemeen kan gesteld worden dat die gegevens a inadequaaf zijn voor wat voor evaluatie ook. zolang ze niet met andere mogelijk meer relevante maar moeilijker te verkrijgen gegevens aangevuld worden. Waar het dan nog gaat om analyse van de zinvolheid van het laten overdoen van tentamens, zitten we bovendien met de moeilijkheid dat doel van de analyse nu juist is om de 'face validity' van het verschijnsel dat scores voor herhaal toetsen meestal inderdaad hoger zijn, op houdbaarheid te toetsen.

De kern van dit argument is, dat de opbrengst van de studietijd besteed aan het overdoen van een tentamen, vergeleken moet worden met de opbrengst van alternatieve studietijd besteding. (er is nog een andere interpretatie mogelijk, de veronderstelling dat het overgedaan hebben de verdere studie zou vergemakkelijken, waarover straks neer.)

Voor de meeste betrokkenen is dit een tamelijk vanzelfsprekende vergelijking, en ik ga er maar van uit dat ook de betrokken docent het eens is met het uitgangspunt dat je geen herhaling van een studieonderdeel moet vragen wanneer een alternatieve studietijdbesteding (gewoon doorgaan met de studie bijvoorbeeld) een hogere opbrengst zou hebben.

Om de uiteenzetting begripsmatig helder te houden, is het mogelijk om, in plaats van te spreken over bepaalde alternatieven, zoals aan het in de eerste bladzijden besproken voorbeeld is gebeurd, te spreken over de te verwachten opbrengst van 'normaal bestede studietijd', en die te vergelijken met de te verwachten opbrengst van het overdoen van een bepaald vak. In het uitgebreid besproken voorbeeld op de bladzijden 79 t/m 84 is wel duidelijk geworden dat het bijzonder onverwacht zou zijn een betere opbrengst voor overdoen dan voor 'normale studie' te vinden. Dat geldt alleen al voor de groep die 'terecht' een onvoldoende had gekregen, en het effect wordt nog versterkt doordat er, zeg, ongeveer evenveel studenten zijn die ten onrechte die

onvoldoende hadden gekregen.

Zo kort kan de toelichting zijn, ik heb er niets meer aan toe te voegen. Maar er zullen wel lezers zijn met bezwaren, en ik denk dat ik een aantal van die bezwaren nu al wel kan noemen.

Tegenwerping 1. Als de student zijn studietijd efficiënt gebruikt, zou hij toch evenveel op moeten kunnen steken als hij een tentamen 'inhaalt', dan wanneer hij die tijd aan andere studieonderdelen besteed zou hebben.

Twee opmerkingen zijn hier van belang: ten eerste dat het moeten overdoen van een studieonderdeel vrijwel altijd extra tijdverlies voor de student met zich mee brengt omdat de normale voortgang van zijn studie er door bedreigd wordt. Ten tweede kost voorbereiding van een herhaal tentamen buiten proportioneel veel tijd, omdat een deel van de al eerder voldoende beheerste stof opnieuw geleerd moet worden, wat afhankelijk van de aard van de stof, en de tijd verstreken sinds de eerste tentamendag, meer of minder tijd zal kosten. Daar komt bij dat het juist in een onderwijs opvatting waarin herhalen van tentamens een rechtmatige plaats wordt toegedacht, er nogal wat voor de student van af hangt of hij die voldoende ook behaalt of niet, met als gevolg dat juist bij de voorbereiding op herhalingen verwacht mag worden dat (althans voor een deel van deze groep studenten) geldt dat zij een stuk meer zullen voorbereiden dan strikt gesproken voor het behalen van een voldoende nodig zou zijn. Tenslotte blijft, ook in een transparante toetsingssituatie, altijd een geluk of pech element aanwezig.

Tenslotte wil ik er nog aan toevoegen dat de tegenwerping voorbij gaat aan de onvermijdelijkheid dat een relatief groot aantal 'herhalers' in feite ten onrechte het tentamen nog een keer moet doen, en voor deze groep geldt per definitie dat iedere tijd aan dit tentamen overdoen besteed, verspilde tijd is.

Tegenwerping 2. De 'alternatieve tijdbesteding' is hier niet relevant, omdat de student die eerst dit tentamen over doet, zijn overige studieactiviteiten gewoon een stukje opschuift; het alsnog voldoende gemaakt hebben levert dan wel degelijk leerwinst op.

Ook hier twee opmerkingen, en om te beginnen al de langzamerhand onvermijdelijke constatering dat deze tegenwerping ook voorbij gaat aan het feit dat een relatief groot aantal van de 'herhalers' het tentamen ten onrechte zit over te doen.

Ten eerste, 'gewoon opschuiven' van de studie komt er op neer dat de afstudeerdatum verlaat wordt. Het valt niet vol te houden dat dat voor de betrokken studenten en voor de samenleving een onverschillige zaak is, en ook de (sub)faculteit zal niet blij zijn met dat soort studieverlenging

Ten tweede verhoogt het overgedaan hebben de overige studieprestaties niet, zodat de prijs voor dat kleine stukje extra stofbeheersing (en dat ook nog alleen voor degenen die het 'terecht' overdeden) te hoog is. Dit raakt echter aan tegenwerping 4., zie daar verder.

Tegenwerping 3. Studenten die het echt niet aankunnen mag je niet zo maar door laten gaan.

Voor zover het gaat om studenten die alleen dit vak niet aankunnen, is het niet duidelijk of grond waarvan geargumentèerd zou mogen worden dat deze student dan maar een fors stuk studievertraging moet op'open, of zelfs dat hij dan ook in deze studierichting niet af zou mogen studeren.

Er zal daarnaast ook een groep studenten zijn die niet alleen met dit vak, maar ook met andere vakken in de studie moeite heeft. Met betrekking tot hen valt op te merken dat, als de (sub)faculteit daar redenen toe ziet, ook zonder dat hen de verplichting opgelegd wordt om allerlei vakken over te doen op grond van de behaalde studieprestaties de raad gegeven kan worden om niet met de studie door te gaan. Dit is het langzamerhand overbekende punt dat niet geprobeerd moet worden om selectie van studenten uit te voeren door middel van het in zijn totaliteit oncontroleerbare zakslaag beleid van de verschillende vakgroepen. Andere manieren, zoals geschetst op blz. 21 en volgende, staan de subfaculteit ter beschikking, als zij al meent op goede gronden nog te kunnen en moeten selecteren onder tot de studie

toegelaten studenten.

Studenten die inderdaad ook met andere vakken moeite hebben de zaakjes laten overdoen, komt er op neer dat zij onder druk gezet worden om de studie af te breken. Het is immers in het geheel niet duidelijk dat het overdoen van een bepaald vak voor hen het doen van de overige vakken in de studie zal vergemakkelijken. Dit is een gedachte waarvan de eerder al geciteerde Jackson (1975) liet zien dat het uitermate moeilijk is om er ondersteunende gegevens voor te vinden. Maar hier komen we al toe aan de volgende tegenwerping:

Tegenwerping 4. Er bestaat een voorwaardelijke relatie tussen bepaalde studieonderdelen, in deze zin dat bepaalde studieonderdelen niet goed gedaan kunnen worden wanneer bepaalde andere studieonderdelen daarvoor niet gedaan zijn (niet op voldoende niveau gedaan zijn.)

Met deze vierde tegenwerping wordt de problematiek iets anders gelegd dan ik tot nu toe besproken heb; de voorspelling is dat het laten overdoen voor de betrokken studenten leidt tot betere studiestatistieken in de verdere studie dan ze zonder het overgedaan hebben behaald zouden hebben. In argument 2 heb ik dat al proberen te weerleggen, en aan het eind van deze paragraaf kom ik er ook uitgebreid op terug.

Is het zo wonderlijk als op het eerste gezicht leek, om als resultaat te vinden dat het laten overdoen niet 'beter' is? Ik denk het niet, en om dat duidelijk te maken mag de volgende toelichting dienen, die anders opgezet is dan het voorafgaande voorbeeld en de daarop aansluitende uitleg.

Als uitgangspunt voor wat voor onderwijsbeleid dan ook, neem ik dat voorzover dat beleid steunt op empirisch toetsbare veronderstellingen, er ook pogingen ondernomen worden om die onderzoekgegevens in handen te krijgen. (dat is eigenlijk ook wat met 'evaluatie van onderwijs, bedoeld wordt: toetsing van de juistheid van veronderstellingen die aan een bepaald beleid ten grondslag liggen, en dan vooral veronderstellingen m.b.t. effecten van een bepaald beleid in vergelijking tot effecten van beleidsalternatieven).

Toegepast op de onderhavige vraagstelling: de veronderstelling van het laten overdoen van tentamens die onvoldoende beoordeeld zijn, is dat de effecten van dit beleid 'beter' zijn dan van denkbare alternatieven zoals het niet laten overdoen, een extra stuk stof laten doen, laten compenseren door cijfers op andere vakken, e.d. De veronderstelling heeft een evident onderzoekbaar of liever: toetsbaar, karakter. Het zou dan ook gepast zijn, om bij onderwijsmaatregelen die zo diep ingrijpen in o.a. de doelmatigheid van het gegeven onderwijs, af en toe die veronderstelling eens kritisch te bekijken en eventueel metterdaad te onderzoeken op haar juistheid. Of het vaak nodig zal zijn om dat onderzoek ook in feite uit te voeren, wil ik betwifelen, omdat het ontwerpen van het experiment, en analyseren van de mogelijke resultaten die het zou kunnen opleveren, al voldoende duidelijk kan maken dat het onwaarschijnlijk is er een bevestiging van de gehanteerde veronderstelling in te vinden.

Stel dat ik wil onderzoeken of over laten doen 'beter' is dan gewoon door laten gaan met de studie, en dat ik 'beter' definieer als 'betere resultaten in de verdere studie' (al dan niet het tijdverlies van het overdoen daarin meerekenend). Stel verder dat ik dat onderzoek wil doen door de helft van de als 'onvoldoende' beoordeelde studenten (random gekozen) het tentamen inderdaad te laten overdoen: de experimentele groep; en de andere helft gewoon door te laten gaan met de studie: de controle groep. (De controle groep bestaat dus uit studenten die een onvoldoende beoordeling kregen, en ontheven worden van de eis om het tentamen over te doen; bij feitelijke uitvoering van een dergelijk experiment zou het nodig kunnen zijn om deze studenten onkundig te houden van de situatie, wat kan door ze mee te delen dat hun tentamen 'voldoende' was). Stel tenslotte dat bij uitvoering van het experiment zou blijken dat de verdere studieresultaten van de experimentele groep beter zijn dan van de controle groep. Volgt daaruit dat over laten doen 'beter' is dan niet over laten doen? Ja. Volgt daaruit dat over laten doen een verstandige maatregel is? Nee. Het laatste is als volgt duidelijk te maken: bij dit gedachte experiment is verzuimd te onderzoeken of het over laten doen misschien ook voor studenten die een voldoende beoordeling kregen, 'beter' zou zijn. Als we dat wel zouden onderzoeken, zou

daaruit kunnen blijken dat ook voor deze studenten geldt dat ze er voordeel van hebben het tentamen nog een keer overgedaan te hebben; een conclusie die vooral makkelijk te verkrijgen is wanneer benodigde studietijd niet meegerekend wordt bij de bepaling van welke studieresultaten 'beter' zijn dan andere.

Maar, als het kennelijk voor iedereen beter is om een tentamen twee keer te doen, dan is er iets niet in orde met de manier waarop we 'beter' gedefinieerd hebben. Een andere mogelijkheid is dat de inrichting van het onderwijs niet louter bepaald kan worden door wat voor een bepaald vak voor de meeste studenten 'betere' resultaten zou opleveren; tenslotte, meer onderwijs is altijd 'beter', maar de totale beschikbare onderwijstijd kan optimaler benut worden door voor de verschillende vakken een tijdbestek beschikbaar te stellen dat lang niet uitputtend is voor wat betreft de mogelijkheden om met de studenten in dat tijdbestek tot 'beter' resultaten te komen. Het maakt in principe niet uit, of die resultaten af te meten zijn aan de directe prestaties voor het betrokken vak, of aan indirecte gevolgen voor verdere prestaties in de studie.

Een variant op de voorgaande foutieve experimentele opzet, is waar effecten van het overdoen worden 'onderzocht' door alleen voor de groep die het overdoet na te gaan of zij 'beter' resultaten bereiken, zeg zoiets als betere resultaten op het betreffende tentamen zelf. Hek', is dan heel erg waarschijnlijk tot conclusies te komen die het kennelijk nut van over laten doen onderstrepen. Dat dergelijke resultaten niets maar dan ook helemaal niets zeggen over het 'nut' van over laten doen is nu wel duidelijk, na alle voorafgaande besprekingen van hoe je dergelijk onderzoek minimaal zou moeten opzetten. Toch wordt dit soort gemankeerd onderzoek zeer veel gepleegd, zoals o.a. op het gebied van onderzoek naar de effecten van het doubleren. Dergelijk onderzoek geeft geen antwoord op de vraag of niet overdoen misschien nog beter is dan overdoen. Of, zoals Jackson (1975) het formuleert m.b.t. doubleerproblematiek:

'It is perfectly possible for grade retention to have some real benefits for retained students, but for grade promotion to have even greater benefits.'

Vervolgens is het ook nog 'perfectly possible' te vinden dat hetzelfde geldt voor studenten die normaal het tentamen niet zouden overdoen: voor hen kun je vinden dat het overdoen goede vruchten afwerpt, maar het niet overdoen nog betere.

Tenslotte is het 'perfectly possible' te vinden dat overdoen voor beide groepen zowel de onvoldoende als de voldoende beoordeelde studenten 'beter' is, in welk geval er in ieder geval geen rechtvaardiging is te vinden voor het beleid dat een deel van de studenten wel en een ander deel het tentamen niet laat overdoen.

Samenvattend kun je concluderen dat voor onderzoek naar de juistheid van veronderstellingen als zou laten overdoen 'beter' zijn dan andere mogelijke maatregelen (gewoon doorgaan met de studie, iets anders er bij doen, compenseren door andere cijfers o.i.d.), tenminste moet gebeuren door na te gaan of:

a. Voor degenen die een onvoldoende beoordeling kregen, geldt dat zij 'betere' resultaten boeken bij het overdoen, dan bij het niet overdoen, en

b. voor degenen die een voldoende beoordeling kregen geldt dat zij betere resultaten boeken bij het niet overdoen, dan wanneer zij (onder experimentele condities) het tentamen zouden overdoen.

Maar dit is precies het onderzoek naar de juistheid van een verondersteld soort verband als aangegeven werd in figuur 5!

De gegeven toelichting vraagt een soort denken dat de lezer niet erg vertrouwd is. Het is dan ook niet erg waarschijnlijk dat het tot nu toe beschreven de overtuiging heeft kunnen geven dat stelling 3. voldoende is toegelicht. Om die zekerheid te verschaffen, is het nodig te laten zien dat de gegeven toelichting ook volledig is, d.w.z. alle argumenten die wel voor het laten overdoen gebruikt worden, ook dekt. Ik zal die argumenten nog eens opsommen, en bij ieder van hen aangeven op welke grond ik meen dat stelling 3 juist is. Hier en daar

zal de tekst herhalingen bevatten van wat eerder in deze paragraaf al aan de orde is geweest.

Aangevoerde bestaansgronden voor het instituut 'overdoen' kunnen zijn:

1. selectie (druk op studenten die relatief vaak vakken moeten overdoen, om de studie te staken);

2. studenten die te weinig tijd besteed hebben aan de voorbereiding, of een gokje wagen, zouden een onvoldoende moeten krijgen zodat ze gedwongen worden om zich alsnog behoorlijk voor te bereiden;

3. de docent wordt in staat gesteld het 'niveau' van zijn vak te beheersen (door overschrijding van de in het studieprogramma in feite voor zijn vak beschikbaar gestelde tijd);

4. als er geen sanctie staat op onvoldoende voorbereiding op het tentamen, zullen de prestaties nog verder zakken (het argument van de lextinsieke motivatie'); en

5. een 'absoluut' minimum niveau moet gehandhaafd worden omdat dat een noodzakelijke voorwaarde is voor het verdere studiesucces, resp. een behoorlijke beroepsuitoefening.

Ad 1. (selectie).

De bedoeling is dan om de student die nogal eens studieonderdelen op deze wijze moet overdoen, onder druk te zetten om de studie te beëindigen (= af te breken). Het is duidelijk dat op deze manier een zeer groot aantal studenten, die uiteindelijk wel af zullen studeren, een stuk studievertraging oplopen omdat andere studenten zo nodig onder druk gezet moesten worden. Maar relevanter is, dat selectie niet via een slinkse omweg gerealiseerd wordt, maar via duidelijke maatregelen geëffectueerd wordt die overigens de kwaliteit van het onderwijs niet nadelig beïnvloeden. Een expliciet selectie beleid maakt het tevens mogelijk om de kwaliteit van die selectie procedure te bewaken, met name ook er voor te zorgen dat een goede balans getroffen wordt

tussen onbedoelde selectie van studenten die anders met voldoende resultaat de studie zouden hebben kunnen beëindigen enerzijds, en onbedoeld door laten gaan met de studie van studenten die alsnog in een latere studiefase stranden anderzijds (zie ook de toelichting op uitgangspunt 5) in het eerste deel van dit rapport).

Een mogelijkheid om selectiebeleid expliciet te maken, is door minimum eisen te stellen aan de totale studieprestaties voor verschillende studieonderdelen gezamenlijk geleverd. Zoals bekend verondersteld mag worden, geeft de gemiddelde studieprestatie over een aantal verschillende studieonderdelen een betrouwbaarder uitgangspunt voor selectie dan een enkel cijfer voor één studieonderdeel. In een soort grade point average stelsel kunnen dergelijke minimum eisen o.a. geformuleerd worden in een minimaal te behalen g.p.a. om bijv. met de vervolgstudie (of bepaalde vervolgstudies) te kunnen beginnen. Dit alles dan onder de veronderstelling dat de subfaculteit zegt te willen selecteren, en daarvoor ook behoorlijke argumenten kan aandragen. Wie niet wil selecteren, kan de selectieve werking van het laten overdoen van studieonderdelen niet als argument voor dat gebruik aanvoeren.

Wat bedoelde selectieve werking betreft kan dan ook geconcludeerd worden dat laten overdoen bijzonder slecht beleid is, en dat daarvoor alternatieven binnen handbereik liggen die de kwaliteit van het onderwijs niet hoeven aan te tasten. Bovendien zijn de directe, en de maatschappelijke, kosten van onnodige studieverlenging direct aan het stelsel van overdoen toe te schrijven. ad 2. (onbehoorlijke voorbereiding.)

Bedoeling van de beoordeling is om iedere student het cijfer te geven dat hij of zij 'verdient', en in dit geval heel in het bijzonder om degenen die zich onbehoorlijk hebben voorbereid in ieder geval een onvoldoende te geven. De idee is erg mooi, maar de kritische vraag is of het in de praktijk mogelijk is om op grond van tentamen scores de studenten die zich onbehoorlijk voorbereid hebben te onderscheiden van studenten die in feite de stof voldoende beheersen maar door de onnauwkeurigheid van de toetsing scores behalen die om de cesuur, of daar beneden liggen. (Voor de duidelijkheid van argumentatie

veronderstel ik hier dat studenten die zich binnen de beschikbare tijd niet voldoende konden voorbereiden, niet aan dit tentamen deelnemen).

Studenten die ondanks onbehoorlijke voorbereiding aan hettentamen deelnemen zijn juist voor een behoorlijke beoordeling erg vervelend. In tabel 3 werd al gedemonstreerd dat naarmate dit aantal studenten relatief groter wordt, er bij de cesuurbepaling steeds meer fouten gemaakt zullen worden. En helaas hebben we door schuiven met de zakslaag grens niet veel keus: scherper stellen van de cesuur resulteert er weliswaar in dat een groot aantal van de studenten met onbehoorlijke voorbereiding een onvoldoende krijgen, maar in hun kielzog slepen zij een groot aantal studenten mee die ten onrechte ook een onvoldoende krijgen. Andersom, het laten zakken van de cesuur, resulteert er in dat een groot aantal studenten met in feite een onbehoorlijke voorbereiding, daarvoor beloond worden door een voldoende beoordeling. Compromissen tussen beide uitersten zijn al even onaantrekkelijk.

Hier blijkt het laten overdoen in geen enkel opzicht een oplossing te bieden voor het probleem dat er mee opgelost zou moeten worden. Er moeten andere maatregelen getroffen worden om het aantal 'malafide' tentamendeelnemers te reduceren (in deel 1 zijn verschillende mogelijkheden genoemd,)

De moeilijkheid is ernstiger dan hierboven beschreven, omdat in de praktijk het aantal studenten dat in feite de stof onvoldoende bestudeerd heeft, niet bekend is. Ook al zou door gedegen onderzoek een soort 'gemiddeld aantal' malafide deelnemers bekend zijn (en de spreiding in dat aantal), dan nog blijft het voor ieder afzonderlijktentamen een gissing hoe groot het aantal malafide deelnemers op dat moment is (zij het dat wel bekend is met welke waarschijnlijkheid dat aantal binnen welke grenzen ligt), zodat het niet mogelijk is om het aantal te geven onvoldoendes af te stemmen op het aantal dat in feite de stof onvoldoende beheerst, met als gevolg dat het aantal foutieve beslissingen opgejaagd wordt. (zie wederom tabel 3.)

Wanneer vermoed wordt dat een relatief groot aantal studenten

met onbehoorlijke voorbereiding aan het tentamen deelneemt, is het al te laat om nog goede maatregelen te kunnen nemen. althans voorzover het gaat om de cesuurbepaling bij dit tentamen. Voor volgende gelegenheden zal naar oplossingen gezocht moeten worden in de sfeer van de inrichting van de onderwijs situatie (voldoende transparantie van toets en cesuurbepaling) en beoordelingsstelsel (een beoordelingsstelsel waarin iedere student ieder tentamen slechts één keer kan doen, zou al te lichtvaardig deelnemen aan tentamens afremmen; inherent aan een dergelijke regeling is natuurlijk dat in de examenregeling er rekening mee gehouden wordt dat incidentele onvoldoendes. door pech behaald, geen belemmering hoeven zijn om te slagen.)

In een stelsel waarin geen herhalingen van tentamens mogelijk zijn, kan de student natuurlijk af en toe ook wel eens een gokje wagen. Het is dan echter makkelijk duidelijk te maken dat dergelijke studie strategie op de lange termijn geen succes kan opleveren. a á n

Een nadeel verbonden grade point average stelsels is dat studenten die al vroeg aan hun minimale verplichten voldaan hebben, en in verband met hun verdere studieplannen geen behoefte hebben aan hogere gemiddelden, hun laatste studieonderdeel (delen) niet of nauwelijks voorbereiden. Het gevolg is dat, als dit verschijnsel zich in enigszins belangrijke vorm voor zou doen, die studieonderdelen geconfronteerd worden met het al gesignaleerde probleem van een relatief groot aantal 'gokkers'. De aanpak van het probleem is in dit geval echter vrij eenvoudig, omdat hier zonder veel pijn de cesuur z6 gesteld kan worden dat zoveel mogelijk studenten die de stof in feite voldoende beheersen, die voldoende ook krijgen. Dat dan ook een groot deel van de 'gokkers' die voldoende in dd wacht slepen heeft geen ernstige consequenties, omdat zij in de eerste plaats toch al aan de minimum exameneisen voldeden, en dat is immers alles waar het hier bij de cesuurbepaling om gaat. Mijn persoonlijke indruk is echter dat niet verwacht moet worden dat in een grade point average stelsel een dergelijk verschijnsel echt omvangrijk zou kunnen zijn: de mate waarin het dat wèl zou zijn, zegt voornamelijk iets over het tekort schieten van de studieprogrammering om de belangstelling van studenten te trekken, of een tekort schieten in relevantie van de studie-inhoud voor eventuele vervolgstudies of het beeld dat de student heeft van de beroepspraktijk. Juist van de 'betere' studenten, zij die al

vroeg een minimaal vereist cijfergemiddelde zullen behalen, is toch te verwachten dat zij ook voor de vakken die zij dan nog moeten doen een meer dan minimale inspanning zullen leveren. (Zie Bevers 1975 voor enkele ervaringen in de Verenigde Staten op dit punt opgedaan).

Dan zijn er natuurlijk altijd nog mogelijkheden om eventuele problemen op te vangen door maatregelen als: naar voren schuiven in het studieprogramma, het vak een groter tijdsbeslag geven, een speciale geavanceerde cursus geven, e.d.

Het handhaven van 'over laten doen' op grond van de overweging dat degenen met een te korte voorbereiding op het tentamen toch geen voldoende mogen krijgen, blijkt niette rechtvaardigen tegenover beleidsalternatieven voor dat laten overdoen. Met name valt een dergelijk beleid niet te rechtvaardigen tegenover die talrijke studenten die er het slachtoffer van worden omdat van hen ten onrechte gevraagd wordt studieonderdelen te herhalen. Dat de student van zichzelf, en de docent van de student niet weet of hij of zij de stof onvoldoende voorbereid had. dan wel gewoon pech had bij het tentamen, doet aan het principiële punt van die rechtvaardigheid niet af: bekend is dat er onder de groep als onvoldoende beoordeelde studenten een groot aantal ten onrechte als zodanig beoordeelden zitten.

ad 3. ('niveaubewaking').

Voorzover de docent het niveau van beheersing van de stof wil bewaken door studenten die in feite onbehoorlijk voorbereid voor het tentamen opkomen, is onder 2 hierboven al beschreven dat daar andere methoden voor moeten worden aangewend. Voorzover het voorkomt dat studenten er van overtuigd Zijn zich behoorlijk op het tentamen voorbereid te hebben, en desondanks onvoldoendes behalen, schort er kennelijk nogal wat aan de voorlichting aan de student m.b.t. aard en inhoud van het tentamen en de wijze waarop de zakslaag grens bepaald wordt (de transparantie van de tentamen situatie is te gebrekkig).

De enig overblijvende categorie studenten die de docent zorgen kan baren omtrent het bereikte 'niveau' zijn degenen die zich behoorlijk hebben ingespannen, maar in de beschikbare tijd de stof niet op het

minimaal vereiste niveau van beheersing konden verwerken. Over het te voeren beleid met betrekking tot deze categorie studenten zal ik het verder hebben.

Een punt van ondergeschikt belang is dat de docent hier misschien teveel uitgaat van een onderwijs doelstelling geformuleerd als een bepaald te bereiken niveau, en te weinig geformuleerd in termen van aanvaardbare of gewenste spreiding van studieprestaties. (Zie het gestelde op blz. 66). Een andere formulering van ongeveer hetzelfde is dat juist voor deze categorie studenten het heel wel denkbaar is dat zij tamelijk veel hebben opgestoken van dit onderwijs (in termen van verschil in stofbeheersing aan het begin en aan het eind van de cursus), maar omdat zij een gebrekkige voorkennis hadden desondanks niet boven het gestelde minimum niveau konden komen binnen de beschikbare studietijd. 'Gebrekkige voorkennis' is natuurlijk maar een relatief begrip, en het wil hier slechts zeggen dat in vergelijking tot andere studenten zij misschien bijzonder gehandicapt waren. Deze situatie doet zich zeer veel vaker in het universitair onderwijs voordan men op het eerste gezicht geneigd is te denken: er zijn nogal wat studies die toegankelijk zijn voor jongens en meisjes met sterk uiteenlopende vooropleidingen, en waar dan ook in het huidige systeem waar al die studenten over één kam gescheerd worden, te zien is dat de onvoldoendes juist bij grote getale vallen in de groep met de vooropleiding die in de onderwijs hiërarchie het laagst genoteerd staat. (voor een recente illustratie van dit overigens algemeen bekende verschijnsel, zie van Vliet 1976).

Nog weer anders geformuleerd: er moet voor gewaakt worden dat doelstellingen van het onderwijs niet ontaarden in alleen maar vaststellen wie bepaalde drempels in de stofbeheersing kan overschrijden, maar dat doelstellingen in meer zuivere vorm gegoten worden als: wat de student in dit onderwijs meer opgestoken heeft, uitgaand boven wat hij aan het begin van de rit al beheerste.

Om de begripsmatige opheldering af te ronden, nog het volgende: Voorzover de docent degenen die binnen de beschikbare tijd de stof niet op voldoende niveau konden beheersen, met de verplichting van overdoen onder druk wil zetten om nog eens te overwegen of

hij of zij wel de juiste studie gekozen heeft, komen we bij het al behandelde punt 1. terecht. Voor zover de docent van mening is dat aan het minimum niveau voldaan moet zijn omdat de student anders grote of onoverkomelijke moeilijkheden in de verdere studie of in de beroepssituatie zal ondervinden, komen we terecht bij het nog te bespreken punt 5. Voorzover de docent dit beleid volgt om de studenten met deze stok achter de deur te dwingen tot grote studie-inspanningen, hebben we te maken met het direct te bespreken punt 4. Het enige overgebleven doel is dan dat de docent gewoon van zijn studenten, althans van een aantal van zijn studenten, een niveau van stofbeheersing vraagt dat uitgaat boven wat in de in het studieprogramma beschikbaar gestelde tijd gerealiseerd kan worden. Kortom, het gaat hier om een regelrechte poging van de student meer tijd te vragen dan in het studieprogramma vastgelegd werd. Daarvoor kan uiteraard geen enkele rechtvaardiging gevonden worden, omdat een dergelijke maatregel allereerst onderwerp van subfacultair overleg hoort te zijn, waaruit eventueel een verzwaring van het betreffende vak, ten koste van andere vakken, kan volgen.

Wie het laatste te bot geformuleerd vindt, kan ik terug verwijzen naar het begin van dit hoofdstuk, waar getoond werd dat voor deze categorie studenten een beter beleid zou zijn in plaats van hetzelfde tentamen nog eens te laten doen, een ander, 'extra' studieonderdeel te laten doen. Hoewel, dat zou een grotere leerwinst opleveren, maar dat zou ook gelden voor de studenten die in de eerste plaats al een voldoende voor het betreffende vak behaalden. Hier wordt mijns inziens het punt bereikt waarop gesteld kan worden dat 'meer' niet 'beter' hoeft te zijn, omdat tenslotte dat 'meer' gekocht wordt tegen een stuk studieduurverlenging. ad 4. (stok achter de deur).

Een beetje primitieve beweegreden voor het handhaven van het instituut 'overdoen' is dat er anders weinig reden meer zou zijn voor de studenten om zich nog maar ergens voor in te spannen. Afgezien daarvan dat ik deze pessimistische visie op de motivatie van de student niet kan delen, is het natuurlijk niet zo dat het verzorgen van onderwijs er uit kan bestaan een aantal strafmaatregelen te hanteren voor degenen die 'uit zich zelf' aan de doelstellingen van het onderwijs willen voldoen. Dat zou een staaltje van zwarte onderwijskunde

zijn, waartegen rationele argumentaties waarschijnlijk weinig kunnen uitrichten.

Bovenstaande opvatting is een karikatuur van het uitgangspunt dat behaalde studieresultaten en beoordeling daarvan enige praktische ('civiele' zo men wil) betekenis moeten hebben. Voor een belangrijk deel moet die betekenisgeving steunen op de inhoudelijke zinvolheid van de gepresenteerde stof. Voor het overige is het meegenomen wanneer het lukt aan de studieprestaties die geleverd zijn ook verder reikende consequenties te verbinden, maar dan vooral in positieve zin (vervanging van examens door dossier diploma's zou een belangrijke stap in die richting kunnen zijn). Maar dit soort beloning voor studiepresties is nog heel wat anders dan het werken met afstraffingen voor 'onvoldoende' prestaties, onvoldoende prestaties bovendien die voor een onbekend deel niet direct de student aangerekend kunnen worden (door feilbaarheid van de toetsing, gebrekkige inrichting van het onderwijs en presentatie van de stof, tekort aan transparantie van de toetsingssituatie, bijzondere persoonlijke omstandigheden, minder goede vooropleiding die echter wel recht op toelating tot deze studie geeft).

Ook al zou het 'stok achter de deur' argument in een bepaald geval toch gehandhaafd worden, dan kan het slechts voor een klein deel van de studenten de beoogde uitwerking hebben, en heeft het misschien voor een ander aantal studenten een ongunstige Uitwerking. Om met de laatste categorie te beginnen: het is niet onwaarschijnlijk dat er onder de groep studenten die wat meer moeite met de studie hebben een relatief niet te verwaarlozen aantal zijn die door de druk die het overdoen systeem op hen zet, te hoog gemotiveerd worden om op het moment van de tentamenafname prestaties te kunnen leveren die representatief zijn voor de mate waarin zij in feite de stof beheersen. (zie het werk van Atkinson c.s. over 'achievement motivation', o.a. Atkinson 1976)..

Voor studenten die geen enkele moeite met deze stof hebben, en voor studenten die er wat meer moeite mee hebben maar een goede studie motivatie hebben, heeft het 'stok achter de deur' effect van het over laten doen bij onvoldoende prestatie op zijn best geen enkele betekenis, en waarschijnlijk zal het in een aantal gevallen interfereren met de

studiestrategie van deze studenten, d.w.z. dat het hen veeleer in de weg staat dan dat zij er zich door aangespoord zouden voelen om enige extra moeite te doen.

Kortom, het 'stok achter de deur' argument gaat voorbij aan de grote spreiding in motivatie, capaciteiten, en studieprestaties binnen vrijwel iedere groep van studenten. En overigens: wanneer studenten het er zonder dreiging van herhalingsoefeningen in grote getale 'bij zouden laten zitten', dan is er waarschijnlijk iets aan de hand met de inrichting, van het onderwijs, de relevantie van de inhoud van het vak zoals die door studenten beoordeeld wordt, of de aard van de toetsing. Dergelijke problemen behoren niet 'opgelost' te worden door de studenten dan maar onder groter tentamendruk te zetten.

Voorzover men van mening is dat er nogal wat 'luie' studenten zijn, die zonder enige aansporing niet tot behoorlijke studie inzet kunnen komen (een opvatting die uiteraard ook voor empirische toetsing toegankelijk is), zou het een opgave voor het te voeren selectie beleid moeten zijn om deze studenten ofwel door goede begeleiding tot behoorlijke studieprestaties te krijgen, ofwel hen via de opgestelde selectie regels de studie uit te loodsen. Dit is op blz. 90 e.v. al behandeld.

Wat 'motivatie' betreft is het laten overdoen op voorhand niet een 'beter' beleid dan het niet laten overdoen, terwijl een fors aantal alternatieven voor handen zijn die wél een behoorlijke motiverende functie kunnen hebben. ad .5. (minimum niveau is voorwaarde voor verder succes).

De redenering is wel bekend: als je dit vak niet voldoende beheerst, zul je met de verdere studie ook moeilijkheden krijgen; als je dit vak niet voldoende beheerst, zul je met bepaalde andere vakken moeilijkheden mogen verwachten; als je dit jaar niet doubleert, zul je zeker het volgende jaar doubleren; als je dit deel van de stof niet begrepen hebt, zul je volgende onderdelen ook niet kunnen begrijpen; als je voor de propedeuse gezakt bent, en je zou gewoon door studeren, zou je zeker ook voor je kandidaats moeten zakken; als je voor de propedeuse gezakt bent, en je zou gewoon doorstuderen, zou je voor

kandidaats of doctoraal niet horen te slagen (variant van De Groot 1972, blz. 107); als je aan dit vak begint, moet je tenminste bepaalde andere studieonderdelen 'voldoende' gemaakt hebben.

Een bepaald minimum niveau als voorwaarde voor verder studiesucces is een wijd verbreide, veel aangevoerde, bestaansgrond voor het instituut 'overdoen'. De weerlegging van deze opvatting belooft ook een boeiend stuk proza op te leveren. Met 'weerlegging' bedoel ik de demonstratie van de empirische onjuistheid van deze opvatting, zo opgevat dat er alternatieven voor het laten overdoen' zijn, die tot betere resultaten, hogere opbrengst of payoff, leiden. Het belangrijkste van die alternatieven is het eenvoudige: niet laten overdoen, natuurlijk ondersteund door andere beleidsmaatregelen omdat het niemands bedoeling kan zijn om van ons onderwijs een laissez faire aangelegenheid te maken.

Voor het volgende wil ik volstaan met eerst nog kort te memoreren dat een relatief altijd belangrijk deel van degenen die een onvoldoende kregen, die onvoldoende ten onrechte kregen, dat een ander deel van degenen die een onvoldoende kregen de studie af zullen maken zonder de onvoldoende overgedaan te hebben, besluiten van studie te veranderen, met de studie op te houden, een ander vak dan dit waar ze de onvoldoende voor kregen te zullen doen, e.d., zodat de minimum, niveau redenering ten hoogste kan gelden voor een heel erg klein aantal studenten. Een uitsplitsing van de te geven toelichting naar verschillende categorieën studenten, zoals in figuur 1 aangegeven, zal niet gemaakt worden; de lezer die daar behoefte aan heeft kan deze oefening zelf maken. Bovendien blijkt de toelichting voldoende te zijn, zonder aan te hoeven geven hoe een en ander geëvalueerd wordt door een bepaalde belanghebbende (groep).

De 'minimum voorwaarde' gedachte, zoals ik dit onderwerp verder zal noemen, laat zich nauwelijks bespreken wanneer niet eerst een aantal kwalificaties worden aangebracht. Gevallen die zich zinvol laten onderscheiden zijn:

- a. het gaat om een tekort aan specifieke kennis, dat door meer van hetzelfde onderwijs bijgespijkerd kan worden;

b. het gaat om een tekort in specifieke capaciteiten, en dat tekort laat zich inhalen door speciaal onderwijs;

c. het gaat om een tekort in algemene capaciteiten, motivatie, hinderlijke persoonlijkheidsvariabelen e.d.

a. Tekort in specifieke kennis, inzichten, vaardigheden.

De opbouw van het studieprogramma naar inhoud is hiërarchisch gedacht, zodat beheersing van voorafgaande onderdelen een voorwaarde is om latere studieonderdelen te kunnen begrijpen. Voor universitaire studies is dat een aanname die na enig nadenken niet direct erg waarschijnlijk is. ik ga daarom bij de behandeling maar uit van een makkelijker voorbeeld: de lezer kan zich voorstellen dat één studieonderdeel opgedeeld wordt in twee delen stof. waarbij beheersing van het eerste deel als voorwaarde wordt gezien om het tweede deel ook maar te kunnen begrijpen. De hiërarchische stof opbouw gedachte zit hier duidelijk genoeg in. Voldoende beheersing van het eerste deel is een voorwaarde voor voldoende beheersing van het tweede deel, althans dat is de veronderstelling. Een veronderstelling die eenvoudig toetsbaar is, bovendien. Nog sterker: het onderzoek naar de houdbaarheid van de veronderstelling is ook noodzakelijk om de aftest grens vast te kunnen stellen, omdat immers de grens tussen voldoende en onvoldoende beheersing in dit geval gedefinieerd is aan het al dan niet voldoende kunnen beheersen (binnen de daarvoor beschikbare tijd) van het tweede studieonderdeel. Voordat ik de haken en ogen inherent aan een dergelijk onderzoekje behandel , wil ik de beschrijving die Salomon (1972) van deze opvatting over het herhalen van stof citeren:

"The assumption made is the same in all: some critical ingredient of knowledge is deficient or missing, and no progress in learning can be expected unless the deficiency is overcome. Consequently, some kind of remedial instruction is called for to close the gap. Underlying this approach is a hierarchical or sequential conception of knowledge, of learning, and consequently also of instruction. It is closely associated with the work of Gagné, Ausubel, and others."

De laatste regel is verrassend: we hadden niet verwacht dat

een populaire opvatting over de functie van het laten overdoen van studieonderdelen verwant is met het tamelijk recente werk van onderwijs psychologen als Gagné (The conditions of learning, 1970) en Ausubel (Educational Psychology 1969). We zullen zien of voorstanders van laten overdoen zich op het werk van beide auteurs kunnen beroepen of niet.

De veronderstelling die gemaakt wordt door degenen die deze rechtvaardiging voor het laten overdoen hanteren, is dat het laten overdoen voor degenen die het eerste deel van de stof onvoldoende beheersten een hogere opbrengst geeft dan het niet overdoen, er bovendien bij implicatie (zie blz. 84) dat laten overdoen voor degenen die de stof wél voldoende beheersten resulteert in een mindere opbrengst dan niet laten overdoen (en aan die laatste voorwaarde is niet zo makkelijk te voldoen. Eerste gedachte is dat voor de laatste groep de tijdbesteding verspild is, en daarom de opbrengst negatief moet zijn. Maar die tijdbesteding is voor de andere groep ook een negatieve opbrengst).

Opbrengst wordt in het hier te bespreken geval afgemeten aan de studieprestaties op het tweede deel, omdat het gaat om een voorspelling van verder studiesucces. Om een interactie zoals geschetst in figuur 5 op blz. 84 te kunnen verwachten moet tenminste aan de volgende condities voldaan zijn (Salomon 1972):

- "a) taskspecific capabilities actually account for a large portion of the variance in the learning outcome,
- b) the learning material is hierarchically or sequentially ordered, and
- c) all the subordinate objectives on the hierarchy are learnable as a function of instruction."

Toelichting bij a): verschillen in studieprestaties op de toets over het tweede deel van de stof moeten voor een groot deel gerelateerd zijn aan verschillen in de mate waarin de stof voor het eerste deel beheerst werd. De tekst van Salomon is daar niet helemaal duidelijk over, maar aan de uitspraak moet in ieder geval toegevoegd worden dat het moet gaan om een oorzakelijk verband tussen beide, en met name niet om

een verband dat ontstaat omdat beide (de resultaten op de toets over het eerste deel van de stof, zowel als Van de toets over het tweede deel van de stof) veroorzaakt worden door verschillen in algemene intelligentie, verschillen in vooropleiding, motivatie, e.d. Het is niet erg waarschijnlijk, of liever: het is erg onwaarschijnlijk om een oorzakelijk verband als hier bedoeld, in het hoger onderwijs tussen verschillende studieonderdelen te kunnen aantonen. Aan het eind van de paragraaf zal besproken worden wat empirisch onderzoek, voor zover bekend, heeft opgeleverd.

Toelichting bij b): eerder verworven kennis en vaardigheid is nodig bij latere verwerving van kennis en vaardigheden. Een dergelijk hiërarchisch verband is in veel situaties nog wel makkelijk te maken wanneer het om de stof binnen één studieonderdeel gaat, maar is veel moeilijker te construeren wanneer het om 'hiërarchische' ordening van complete studieonderdelen gaat. Voor universitaire studies geldt nu eenmaal in hoge mate hetzelfde wat ook al voor verschillende vakken in het lesprogramma voor lager of middelbaar onderwijs geldt: de hiërarchische relaties tussen verschillende vakken zijn er vrijwel niet, en binnen eenzelfde vak over verschillende leerjaren heen ook maar in beperkte mate.

Toelichting bij c): omdat de remedie wordt gezocht in het laten volgen van nog eens hetzelfde stukje onderwijs, is de onderliggende gedachte kennelijk dat dat onderwijs de bepalende factor is voor de mate van beheersing van die specifieke kennis en vaardigheden. Dat is in tegenstelling tot specifieke capaciteiten die slechts met bijzondere onderwijsprogramma's op te vijzelen zijn, of zelfs helemaal niet, en algemene intelligentie bijv. die vrijwel niet door onderwijs beïnvloedbaar is)

Verdere argumentatie voor de juistheid van min stelling 3. voor het onderhavige geval hoef ik niet te geven, gezien het hele verhaal op blz. 71 t/m 78 te vinden is ! Daarom als overgang naar het volgende punt dit citaat uit Salomon (1972):

"Moreover, in certain learning tasks, general psychological states (e.g., anxiety) play a major role in distinguishing between the

successful and the less successful learner. The remedial model does not seem to apply to these cases either. Providing highly anxious learners with remedial instruction would not affect the cause of their failure i.e., it would not lead to much improvement."

Een onderzoek methode en resultaten van een eerste toepassing daarvan worden gegeven door Fryor & Gordon, toegespitst op de onderwijs situatie waar voor een aantal hoofdvakken verondersteld wordt dat een aantal andere vakken als voorbereiding daarop tenminste gedaan moeten zijn. In hun inleiding kunnen de auteurs het niet laten de volgende opmerking te maken:

"Decisions about the necessity and scope of course prerequisites generally rely upon anecdotal evidence, personal experiences, and common sense. Fortunately, more objective procedures exist for resolving the oftentimes hotly debated matter of prerequisites. Unfortunately, however, these empirical approaches appear to have been almost completely overlooked by educators."

Vervolgens schetsen zij hun methode, die vooral bruikbaar is in het geval het in de praktijk toch wel voorkomt dat studenten aan hoofdvakken beginnen zonder alle vereiste voorbereidende studieonderdelen gedaan te hebben. De methodiek bestaat slechts uit een bijzonder moeilijk inzichtelijke statistische manipulatie van beschikbare gegevens. Beschikbare gegevens, d.w.z. dat niet gewerkt wordt met experimenteel gecontroleerde condities, wat een aantal methodologische bezwaren heeft (de auteurs besteden aan dit probleem geen enkele aandacht, helaas. Het is niet duidelijk in hoeverre de interpreteerbaarheid van hun resultaten er door verzwakt wordt; het is mogelijk dat bij veranderd beleid, wat zij ook voorstellen, andere resultaten verkregen zouden worden dan zij hier rapporteren).

Het gaat hier om de vraag of studiesucces op de hoofdvakken in belangrijke mate bepaald wordt door het al dan niet gedaan hebben van de vereiste voorbereidende vakken (dat is conditie a) genoemd door Salomon, zie blz. 97). Welnu, zij vonden:

"Grade point average was found to be the variable most closely

associated with the level of performance in the target courses, whereas the percentage of prerequisites accounted for approximately two percent of the variance."

Met andere woorden: ook wanneer leerstof van verschillende vakken als hiërarchisch geordend wordt beschouwd, d.w.z. het ene vak is een voorwaarde voor het. behoorlijk kunnen doen voor het andere vak, blijkt dat studieresultaten op het laatste vak vrijwel niets te maken hebben met het al of niet gedaan hebben van die voorbereidende vakken, maar wel met het gemiddelde cijfer tot nog toe voor de studie behaald. Ofwel: niet het al dan niet gedaan hebben van een voorbereidend vak is bepalend voor studiesucces, maar het beschikken over die eigenschappen als algemene intelligentie, motivatie, e.d. welke eerder studiesucces in de eerste plaats mogelijk maakten, bepalen ook verder studiesucces in hoge mate. Het laatste is mijn interpretatie, en niet die van Pryor & Gordon.

De opmerking van Pryor en Gordon dat onderzoek over het nut van voorbereidende vakken ontbreekt, is echter aantoonbaar onjuist. Zo is er een bekend onderzoek van Ahmann & Glock (1959), waarin het effect van een voorbereidende cursus wiskunde voor eerste jaars college studenten, met gebrekkige voorkennis van wiskunde, op verder studiesucces onderzocht werd. De samenvatting van dat onderzoek:

"The effectiveness of an experimental course in applied mathematics for agriculture freshmen (except those whose mathematical skill is superior) was evaluated by comparing an experimental and a control group in terms of subsequent grade point averages, final marks in courses involving mathematics, gains in knowledge of mathematics, and tendency to remain enrolled at Cornell.

In addition, the opinions of the students enrolled in the course and their faculty advisers concerning the course were obtained.

The statistical evidence does not indicate that the experimental course influenced subsequent achievement or ultimate attrition rate to any important degree. However the student and faculty opinions concerning the worth of the course were favorable."

De indruk van alle betrokkenen was ook hier in tegenstelling tot het feitelijke effect van de betrokken cursus: dat ze niets bijdraagt, en in feite een schadelijke zaak is omdat het volgen van deze cursus de studenten nu eenmaal wel tijd gekost heeft. (De controlegroup kreeg de cursus niet, en behaalde dezelfde verdere studieresultaten. Merk op dat studenten met 'superior mathematical skill' niet in het experiment werden betrokken, wat bij een goede opzet van onderzoek naar de juistheid van een beleid die verschillende studenten aan verschillende onderwijs behandelingen toedeelt, wel nodig zou zijn. Achteraf bezien, is het niet jammer dat zij deze groep studenten niet in hun onderzoek betrokken hadden, omdat resultaten die dat extra zou opleveren niets meer zouden kunnen veranderen aan de conclusie dat de extra cursus overbodig was).

Een andere, voor ons onderwerp interessante, studie wordt door Hills (1970) geciteerd:

"An unpublished study by Gelso and Trowell (1967) carried the placement question through to a reasonable criterion and discovered that the plan for placement in history at South Georgia College, a public junior college in Douglas, if anything, reverses the order of the courses. In this instance, the courses History 110 and History 111 are considered by the faculty to be logically sequential though it is recognized that the students' understanding of the material in History 111 is not dependent on their understanding of the material in History 110. The question that was raised was whether students would obtain higher grades in 111 if it was taken after 110. Of two groups of students, one took the courses in the logically sequential order (110 followed by 111), and one took the courses in the reverse order (111 followed by 110). Instead of the logical order producing higher grades in the second course, the results indicated that there were no significant differences, but the mean grades were higher in both 110 and 111 if the courses were taken in the logically incorrect order, 111 followed by 110. Although Gelso and Trowell did determine that the groups were equivalent on predicted freshman average grade, they were not able to assign students randomly to sequences. There might then be relevant variables operating that were not controlled by finding only

minute differences between groups in predicted averages. At any rate, the study shows the importance of checking on a placement procedure that seems to be logically sound."

b. Tekort in specifieke capaciteiten.

Ik verwacht voor universitair onderwijs in het algemeen niet dat verschillen in specifieke capaciteiten belangrijk genoeg kunnen zijn om in grote mate verschillen in beheersing van de stof voor haar rekening te kunnen nemen. Een enkele uitzondering valt te maken, bijvoorbeeld voor zoiets als helt onderwijs in de tandheelkunde waar vastheid van de hand een belangrijke specifieke eigenschap is. Voor zover een dergelijke eigenschap in oorzakelijke relatie staat tot later studiesucces, zoals bijv. beoordeling van practicum werkstukken, is het evident dat laten overdoen van studieonderdelen bar weinig zal kunnen helpen.

Hier moet iets aan de specifieke capaciteit zelf gedaan worden (door onderwijs dan wel door therapeutische maatregelen, en als dat niet mogelijk blijkt, en verder studie of beroeps succes zou daardoor in gevaar komen, dan kan nog overwogen worden om het verdere onderwijs minder afhankelijk te maken van die specifieke vaardigheid (hetzelfde geldt voor de beroepsuitoefening) door het maken van de nodige extra onderwijs hulpmiddelen, óf er moet op die specifieke eigenschap een toelatingsselectie gebaseerd worden. In ieder geval kan dit met overdoen van studieonderdelen niets te maken hebben.

c. tekort in algemene capaciteiten, etc.

Het is duidelijk dat onvoldoende studieresultaten op een bepaald vak die terug te voeren zijn op algemene capaciteiten van de studenten, en op 'psychological states' zoals Salomon dat noemt, waarschijnlijk wel degelijk voorspellende waarde hebben ten aanzien van verder studiesucces, maar alleen vanwege die achterliggende algemene eigenschappen als intelligentie, motivatie, e.d. Even duidelijk moet dan ook zijn dat het laten overdoen van een bepaald studieonderdeel aan de voorspelling niets zal veranderen, omdat die algemene intelligentie, motivatie etc. daardoor niet veranderen! Overgedaan hebben van een bepaald vak bereidt deze studenten dan

ook niet beter voor op de rest van de studie, behalve misschien doordat zij intussen wat ouder,'en daardoor misschien wat fwijzer' zijn geworden. Hetzelfde probleem dus als met het doubleren in het lager onderwijs: relatief slechte studieresultaten in het ver volgonderwijs worden daarmee niet voorkomen (zie de gegeven citaten van Jackson in het voorgaande). Het kan nimmer de bedoeling zijn om door het laten herhalen van studieonderdelen de algemene intelligentie van de student op te voeren, zijn tentamenvrees te verminderen, motivatie te veranderen (voorzover die motivatie betrekking heeft op diepliggende persoonlijkheidseigenschappen en niet op de manier waarop het onderwijs gepresenteerd wordt). Afgezien van de feitelijke onmogelijkheid daarvan zou zoiets ook niet in de doeistellingen van onderwijs thuishoren. Waar het dan uiteindelijk op neer komt is dat studenten met relatief geringe intellectuele of motivationele etc, bagage, hij voorkeur uit worden geselecteerd door het stelsel van over laten doen. Dit is een oud en bekend punt: als men wil selecteren, en daar zijn goede argumenten voor aangevoerd waarin de betrokkenen zich kunnen vinden, dan moet die selectie ook heel expliciet zijn, en niet in het stiekeme en oncontroleerbaar gebeuren door een opeenvolging van gedwongen herhalingen van studieonderdelen.

Ik wil de toelichting op argument 3. hier maar afsluiten. Te overwegen valt nog of ik in de definitieve versie niet een bespreking moet invoegen van de onderzoeken (Pryor & Gordon, Hills) die negatieve resultaten opleverden, met name ook voor de 'minimum voorwaarde' gedachte. Met gekoppeld daaraan nog een bespreking waarom het niet verwonderlijk is om dergelijke resultaten te vinden (wat op blz 71-78 ook al besproken is)

LITERATUUR.

Ahmann, J. S., & Glock, M. D. An evaluation of the effectiveness of a freshman mathematics course. *Journal of Educational Psychology*, 1959 50, 4145. Ook besproken in Hills, 1971.

Alfj E. F.9 & Dorfman, D. D. The classification of individuals into two criterion groups on the basis of a discontinuous payoff function.

Psychometrika, 1967, 321 115123

Atkinson J. W., Lens, W., & O'Malley P. M. Motivation and ability: Interactive psychological determinants of intellectual performance, educational achievement, and each other. In Sewell, W. H., Hauser, R. M., & Featherman, D. L. (Editors) *Schooling and achievement in American society*. London: Academic Press, 1976.

Beyers, J. A. A. M.: Het Amerikaanse studiepuntenstelsel. Een literatuurstudie. In: Commissie Ontwikkeling Hoger Onderwijs: De invoering van een studiepuntenstelsel in het hoger onderwijs. Den Haag: Staatsuitgeverij, juni 1975

Bishop" Y. M. M., Fienberg, S. E., & Holland, P. W.: *Discrete multivariate analysis, theory and practice*. London: M.I.T. press, 1975.

Bloom, B. S.: *Stability and change in human characteristics*. London: Wiley, 1964.

Bloom, B. S., Hastings, J. T., & Madaus, G. F.: *Handbook on formative and summative evaluation of student learning*. London: McGrawHill, 1971.

Dawes. R. M.: Graduate admission variables and future success. *Science*, 1975, 187, 721-723

Groot, A. D. de, Waaraan voldoet een 'onvoldoende' prestatie niet? *Paedagogische Studiën*, 1964, 41, 116.

Groot, A. D. de : De kernitemmethode voor de bepaling van de caesuur voldoende/onvoldoende. *Paedagogische Studiën*, 1964 41, 425-440.

Groot, A. D. de : Vijven en zessen. Cijfers en beslissingen: het selectieproces in ons onderwijs. Groningen: Wolters, 1966.

Groot, A. D. de *Selectie voor en in het hoger onderwijs. Een probleemanalyse*. Den Haag: Staatsuitgeverij, 1972.

Hillsj J, R. : Use of measurement in selection and placement. In: Thorndike. R. L. (Editor) Educational measurement. Washington, D. C.: American Council on Education, 1971.

Hofstee, Yv. K. B. : De betrouwbaarheid van zak-slaagbeslissingen. Nederlands Tijdschrift voor de Psychologie, 1970, 25, 380-383.

Hofstee, W. K. B. : Participatiekontrolle door 'onbenullige* toetsitems. Nederlands Tijdschrift voor de Psychologie, 1973, 28, 189-198.

Jackson, G. B. : The research evidence on the effects of grade retention.

Review of Educational Research, 1975, 4, 613-636

Lindley, D. : Making decisions. London: Wiley, 1971.

Lord, F. M., & Novick, M. R. : Statistical theories of mental test scores. London: AddisonWesley, 1968.

Naerssen, R. F. van : Psychometrische aspecten van de kernitemmethode. Nederlands Tijdschrift voor de Psychologie, 1974, 421-430.

Novick, M. R., & Jackson, P. H. : Statistical methods for educational and psychological research. London: McGraw Hill, 1974.

Page, E. B., Jarjoura D., & Konopka, G. D. : Curriculum design through Operations Research. American Educational Research Journal, 1976, 13, 31-50.

Pitz, G. F. A structural theory of uncertain knowledge. In Wendt, D., & Vlek, Ch. (Editors) Utility, probability and human decision making. Dordrecht; Reidel, 1975.

Pryor, N. M., & Gordon, M. E. : A statistical model for the examination of the validity of prerequisites. Educational and Psychological Measurement, 1974, 141, 349-356.

Raiffa, H. : Decision analysis. Introductory lectures on choices under uncertainty. London: AddisonWesley, 1968.

Salomon, G. : Heuristic models for the generation of aptitude treatment interaction hypotheses. Review of Educational Research, 1974, 42, 327-344.

Standards for educational and psychological tests. Washington, D. C., 1200 Seventeenth Street, N. W.: American Psychological Association, 1974.

Thorndike, R. L. (Editor) Educational measurement. Washington, D. C.: American Council on Education, 1971.

Walberg, H, J, : Social environment and individual learning: a test of the Bloom model. Journal of Educational Psychology, 1972, 63, 6973

Wijnen, W. H. F. W. : Onder of boven de maat. Een methode voor het bepalen van de grens voldoende/onvoldoende bij studietoetsen. Amsterdam: Swets & Zeitlinger, 1972.

overdoen afschaffen. 24976.

Voorlopige samenvatting t.b.v. discussie vrijdag 24 september.

Onderwerpen zijn de wijze waarop cesuurbepaling plaats moet vinden, en de wenselijkheid om het laten overdoen van 'onvoldoende' studieonderdelen af te schaffen.

Beide onderwerpen zijn niet los van elkaar te zien:

a) de mate waarin overdoen afgeschaft is, maakt de cesuurbepaling veel minder kritisch (en problematisch)

b) een behoorlijke opzet van de cesuurbepaling resulteert in een nog maar gering aantal onvoldoendes (o.a. omdat behoorlijke

cesuurbepaling een aantal onderwijs verbeterende maatregelen impliceert).

Voor een goed begrip: met 'overdoen' wordt altijd bedoeld dat de student opnieuw tentamen af moet leggen over de stof, en wel zo dat hij daar erg veel tijd aan kwijt is omdat hij inderdaad de hele stof opnieuw moet bestuderen o.a. omdat hij tegen de tijd van de volgende tentamengelegenheid anders een groot deel vergeten zou zijn. Herkansingen die de volgende dag bijv. gegeven worden vallen daar zeker niet onder.

Belangrijk praktisch verschil tussen beide onderwerpen is dat de wijze van cesuurbepaling voornamelijk de verantwoordelijkheid van de individuele docent (vakgroep) is, terwijl afschaffen van het overdoen tenminste een subfaculteitsmaatregel is. Op grond van de gegeven uiteenzetting kan de individuele docent (vakgroep) onmiddellijk een aantal maatregelen nemen die op de cesuurbepaling betrekking hebben. De noodzaak daartoe valt bovendien makkelijker in te zien op grond van de argumenten die betrekking hebben op de overbodigheid, zinloosheid, en schadelijkheid van de gewoonte onvoldoendes te laten overdoen.

In vergelijking met bekende cesuurbepalingsmethoden (kernitemmethoden, methode Wijnen) onderscheidt de hier geboden benadering zich op de volgende punten:

a) een aantal onderwijsmaatregelen zijn tenminste nodig om het cesuurprobleem beheersbaar te maken;

b) de methode wordt geconstrueerd in overeenstemming met een aantal uitgangspunten van normatieve, ethische, en juridische aard; uitgangspunten waartegen overigens iedere methode voor cesuurbepaling getoetst moet worden. Dergelijke toetsing, ook wat betreft onderwijskundige uitgangspunten die onder a) thuishoren, valt voor de kernitemmethoden vrijwel onveranderlijk zeer slecht uit, terwijl de methode Wijnen aanvaardbaar is als oplossing op de korte termijn.

Het is niet de bedoeling geweest in dit rapport de methode in gedetailleerde vorm te presenteren. Het rapport kan dan ook niet dienen als handleiding voor het uitvoeren van de voorgestelde wijze van cesuurbepaling. Desondanks zijn een aantal beleidsaanbevelingen voor de docent (vakgroep) op basis van het behandelde wel te geven, voornamelijk is dat om tenminste de methode Wijnen te hanteren. (zie blz. 50).

Omdat in de dagelijkse onderwijs praktijk de cesuurbepaling nauw verweven is met het selectie mechanisme, wordt op verschillende plaatsen aan het onderwerp selectie aandacht besteed (vooral blz. 22-28). Daarin wordt er de nadruk op gelegd dat dergelijke selectie kennelijk plaats vindt, en dat men er niet voor terug mag deinzen het onderwerp expliciet te bespreken (uit de taboesfeer te halen) en het faculteitsbeleid op het punt van selectie geen enkele onduidelijkheid te laten bevatten. Selectie, als er geselecteerd wordt, is een zaak waar de subfaculteit ook alle verantwoordelijkheid voor behoort te dragen (en niet de individuele docent of vakgroep). Selectie en cesuurbepaling horen dan ook niets met elkaar te maken te hebben. Géén selectie via cesuurbepaling, dat is een eerste voorwaarde voor behoorlijke cesuurbepaling.

Hoewel dat niet direct in de macht van de docent (of vakgroep) ligt, wordt er telkens weer op gewezen dat overgang naar een geheel ander beoordelingsstelsel, een stelsel als het Amerikaanse grade point average systeem, bijzonder gewenst is. Aanvaarding van een dergelijk stelsel zou het vanzelfsprekend maken dat overdoen afgeschaft wordt.

Dat afschaffing van het overdoen een dringende zaak is, wordt in het tweede deel van het rapport aan de hand van een aantal verschillende argumentaties, aangetoond. Een grote rol speelt daarbij het feit dat altijd, bij welke cesuurbepaling dan ook, een relatief belangrijk aantal foute beslissingen worden gemaakt. Wanneer men bovendien er aan 'gewend' is forse afwijzingspercentages te hanteren, betekent dat dat grote aantallen studenten ten onrechte onvoldoendes krijgen, en daardoor studievertraging oplopen. Opmerkelijk is dat docenten zich onvoldoende bewust zijn van de omvang van dat aantal foutieve beoordelingen, terwijl daar tevens de belangrijkste reden voor

de vaak volstrekke zinloosheid van het laten overdoen in gezocht moet worden.

Het blijkt dat afschaffing van het overdoen een beleidsmaatregel is die tamelijk eenvoudig door te voeren valt, niet resulteert in vermindering van de onderwijskwaliteit ofwel de kwaliteit van de afstuderenden, maar wel resulteert in een grote gemiddelde studieduurverkorting. Het laatste is economisch gezien van belang, omdat de besparing voor de onderwijsbegroting aanzienlijk is, en er bovendien een niet verwaarloosbaar economisch effect aan verbonden is dat zich niet direct in begrotingscijfers uitdrukt (vervroegde maatschappelijke intrede van academici). Het voordeel van een sterke studieduurverkorting kan uiteraard ook gebruikt worden om de studie inhoudelijk te verrijken, zodat de studieduur gelijk blijft, maar de academische bagage van de afstudeerders vergroot is.

Het laat zich aanzien dat in de gedane voorstellen misschien één van de zeldzame mogelijkheden ligt om de herstructureringsoperatie te laten slagen, althans op het punt van het nagestreefde studierendement binnen de daarvoor gedachte studietijd.

overdoen afschaffen. wat nog meer?

Het zou mooi zijn het bij 100 blz. te kunnen laten, maar ik ben nog niet helemaal rond met mijn verhaal. Wat komt er nog meer:

1. een toelichting op de overige stellingen genoemd op blz. 3 en 4, toelichtingen die naar ik verwacht echter niet langer hoeven te worden dan door de bank genomen één A4 per stelling.

2. Als bijlage een korte uiteenzetting van de kernitemmethode De Groot, de herziene, kernitemmethode Van Naerssen, en de methode Wijnen.

3. Een bijlage waarschijnlijk verzorgd door Michel [Korzac] over conclusies op basis van dit rapport m.b.t. de voorstellen over invoering van een studiepuntenstelsel.

4. Een beschouwing over de problematiek van compenseren of niet (hoewel dat niets meer toevoegt aan wat ik in de laatste bladzijden, 96 tot 100, beschreven heb.

Een aardige samenvatting.

6. Zo mogelijk een aantal discussiepunten voor de docentlezer.

7. Een aantal beleidsconclusies aansluitend op de toelichtingen op de stellingen.

8. De bibliografie dan natuurlijk nog.