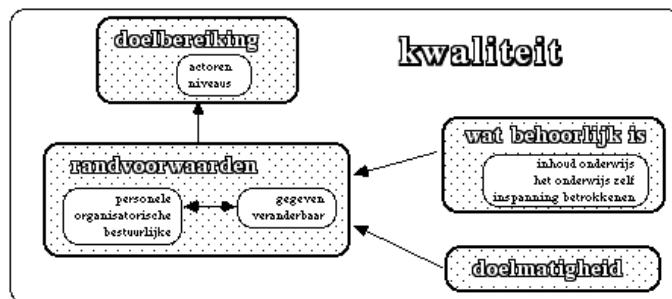


Efficiënt toetsen.

Ben Wilbrink, SCO/UvA

Voordracht HBO-congres over extensivering. 18 februari 1987 (Utrecht) en 24 februari (Eindhoven). (organisatie: Carlo Hover)

Deze voordracht gaat over de mogelijkheden om het beoordelingsproces als hefboom te gebruiken om de kwaliteit van het onderwijs te verbeteren, ik zal daarom eerst de relatie van extensivering tot onderwijskwaliteit aanduiden.



Figuur 1. Wat is kwaliteit in het onderwijs? (B. Wilbrink en Th. Joostens Kwaliteitsbewaking W.O. Amsterdam: S.C.O. / Sectie WO., dec. 1986)

Extensiveren is bezuinigen op de middelen zonder de kwaliteit aan te tasten, dus binnen knappere randvoorwaarden dezelfde prestatie leveren. Extensiveren is alleen mogelijk door doelmatiger met de resterende middelen te werken, dus door organisatorische en onderwijskundige ingrepen die op zich kwaliteitsverhogend zouden werken. Maar als dat zo is, dan moet het zo zijn dat alle maatregelen die een positief effect op de kwaliteit van het onderwijs hebben, in beginsel in een extensiveringsbeleid kunnen worden ingezet. Laten we dan nu spreken over de kwaliteit van het onderwijs in relatie tot de wijze waarop er in het onderwijs wordt getoetst en geëxamineerd. Mijn eerste opmerking moet dan zijn dat er op het gebied van het beoordelen heel weinig onderzoek wordt gedaan: er is een groot instituut dat zich met toetsen bezighoudt, het CITO, maar daar gebeurt geen onderzoek naar de wijze waarop onderwijs en beoordeling zich tot elkaar zouden moeten verhouden. In de zestiger jaren is er een 'Bureau Examenteknieken' geweest, waar Adriaan de Groot en Bob van Naerssen baanbrekend werk hebben gedaan, naast het propageren van de objectieve studietoets in een land waar tot dan toe bij voorkeur mondeling werd geëxamineerd. Aan de Universiteit Twente wordt nogal theoretisch onderzoek verricht, en daarmee is, op enkele individuele inspanningen na,

het onderzoek in ons land eigenlijk wel aangegeven. Het is misschien wel een bijzondere gelegenheid die u nu wordt geboden: in het bestek van een half uur enig inzicht te verwerven over de wijze waarop het gedrag van studenten is gerelateerd aan de wijze waarop er wordt getoetst, en hoe in principe door in te grijpen in de wijze van toetsen dat gedrag van de studenten kan worden beïnvloed, ten gunste van doorstroming en numeriek rendement, ten gunste van de kwaliteit van het onderwijs dus, en daarmee ook mogelijkheden voor extensivering biedend.

Over cijfers, over slaagpercentages, over het aantal toegestane herkansingen, en tal van andere onderwerpen kunnen docenten heftig met elkaar van mening verschillen, en hebt u zelf waarschijnlijk ook zeer uitgesproken opvattingen. Ik wil laten zien dat dergelijke verschillen van mening in principe kunnen worden beslecht met objectieve argumenten, en ten gunste van de kwaliteit van het onderwijs.

over toetsen in het onderwijs

Onderwijsverbetering is in deze eeuw een moeizame zaak gebleken. Wie zoekt naar mogelijkheden voor verbetering, ook al is dat slechts het behoud van kwaliteit bij vermindering van in te zetten middelen, staat voor de vraag welk soort ingreep tot interessante effecten kan leiden.

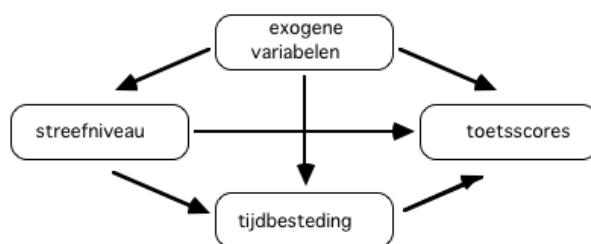
Het onderwijs veranderen is niet moeilijk, maar bij zowel grootschalige als kleinschalige projecten blijken er onverwachte neveneffecten op te treden die een deel van de beoogde winst tenietdoen. De eersten die van verbeteringen in het onderwijs zullen profiteren zijn de studenten, en dat laten zij blijken door tegen een nu geringere inspanning dezelfde resultaten als voorheen te behalen. Dat is een ervaringsfeit dat niet tot treurnis hoeft te stemmen: het wijst juist de weg naar een aanpak die wél tot resultaten zal leiden. Het gedrag van de studenten is bepalend voor wat in het onderwijs kan worden bereikt, en het gedrag van de studenten wordt gestuurd door de wijze waarop er wordt getoetst en de eisen die worden gesteld. Een Individueel StudieSysteem kent een intensieve en frequente toetsing, met sterke effecten op de inspanning van de studenten. Een belangrijk deel van die kracht komt voort uit de verbeterde terugkoppeling die de studenten over hun vorderingen krijgen. Meer algemeen geldt dat het beoordelingsstelsel het krachtigste instrument is om veranderingen in het onderwijs te bewerkstelligen, zoals het ook de sterkste behoudende kracht in het onderwijs vormt, mits ongemoeid gelaten. In 1940 demonstreerde Posthumus in een artikel in De Gids dat het beoordelingsstelsel in het voortgezet onderwijs steevast tot jaarlijkse afwijzingspercentages van rond de 25 % had geleid, over alle vergaande maatschappelijke ontwikkelingen van de voorafgaande eeuw heen. Tot op de dag van vandaag is het onbekend wat de aard is van het ongetwijfeld complexe mechanisme achter dit verschijnsel, maar men mag veronderstellen dat het zowel het beoordelingsgedrag van docenten is, als de wijze waarop leerlingen reageren op deze beoordelingsgewoonten, die tot de wetmatigheid van Posthumus leiden. Binnen het beoordelingsstelsel is het de **wisselwerking** tussen gedragingen van docenten en studenten

die de onderwijsuitkomsten bepaalt. De eerste uitwerking van deze gedachte is in 1970 gegeven door Van Naerssen in de vorm van een wiskundig model voor de minimale tijd die nodig is om voor een tentamen te slagen, inclusief eventueel nodige herkansingen. Een andere belangrijk inzicht werd omstreeks dezelfde tijd verwoord door Adriaan de Groot, waar hij het belang onderstreepte van doorzichtigheid van de toetssituatie: de student moet in staat worden gesteld zich adequaat voor te bereiden op een examen. Doorzichtigheid is een geenszins vanzelfsprekend concept: het staat haaks op de geheimzinnigheid die juist bij psychologische tests wordt gehandhaafd, en in misplaatste navolging daarvan soms ook bij toetsen in het onderwijs.

Ik zal eerst het gebruik van toetsen en examens als regelmechanisme uiteenzetten; daarna wil ik ingaan op de kwaliteit van toetsen, en zal ik mij wagen aan enkele praktische aanbevelingen voor kwaliteitsverbetering, of extensivering, zo u wilt.

toetsen als regelmechanisme

Het gaat de student erom zich te kwalificeren als ingenieur of leraar, econoom etcetera. Het gaat de student om het behalen van het examen, dus om het slagen voor de onderdelen van dat examen. Een examen is daarmee veeleer een eindstreep dan een meetinstrument. Het verschil is niet onbelangrijk, omdat betrokkenen op alle niveaus examens en toetsen meestal behandelen alsof het meetinstrumenten zijn, en dat is een kostbare misvatting. Gooi geen geld weg aan het verbeteren van het examen in meettechnische zin, maar maak de weg voor de student naar die eindstreep toe recht, haal obstakels weg, en breng markering en verlichting aan. Zorg ervoor dat studenten zich doelmatig op toetsen en examens kunnen voorbereiden.



Figuur 2. Studieresultatenmodel.

Studenten proberen hun doelen op de meest efficiënte wijze te bereiken. De figuur geeft een beeld van de ingrediënten van het gedrag van de student. Wat hij of zij meebrengt zijn de exogene variabelen, daartoe behoort de genoten vooropleiding, en de intellectuele en andere capaciteiten waarover men beschikt. De student kiest zich een

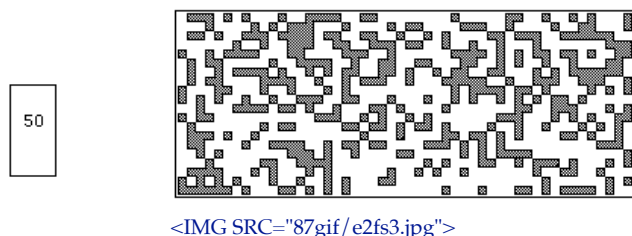
streefniveau, een bepaalde kwaliteit van de eigen prestaties in vergelijking tot die van anderen of in vergelijking tot meer absolute of idealistische normen. De student investeert tijd in de voorbereiding op toets of examen. De pijlen geven oorzakelijke verbanden aan, merk op dat de pijlen een bepaald patroon vormen. Ik gebruik dit beeld graag als steun bij het denken over wat er in het complexe onderwijsveld, het grijze gebied, gebeurt.

Als de pijlen oorzakelijke verbanden aangeven, hoe is dan de invloed van het beleid op het gedrag van de student voor te stellen? Het aardige is dat de beleidsmogelijkheden zo talrijk zijn dat ze niet even eenvoudig in de figuur zijn mee te nemen. U kunt het onderwijs aantrekkelijker maken, zodat de student beter gemotiveerd zich een hoger streefniveau kiest. Doorzichtige toetsing leidt tot doelmatige tijdbesteding, wat voor sommige studenten minder tijd is, voor anderen mogelijk meer tijd omdat ze nu inzien dat de tijd goed besteed is. Een belangrijk beleidsinstrument ligt in de eisen die men aan de student stelt, bijvoorbeeld waar het gaat om zakken en slagen voor toetsen of examens.

aftestgrenzen.

Er bestaan zeer veel misverstanden over het beoordelen in het grensgebied tussen 'voldoende' en 'onvoldoende'. De problematiek kan echter tamelijk helder als volgt worden beschreven.

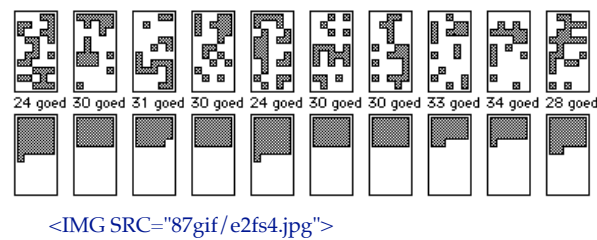
Iedere toets is een steekproef uit een verzameling van toepasselijke toetsvragen. Een dergelijke verzameling kan als volgt worden voorgesteld:



Figuur 3. Verzameling van 1000 toetsvragen ('itembank'). (B. Wilbrink Handleiding zelf-evaluatie Rechtsgeleerdheid 1986-1987. Amsterdam: S.C.O./Sectie H.O., november 1986)

In de grote rechthoek zijn 1000 hokjes weergegeven, die ieder een toetsvraag representeren. Een denkbeeldige student heeft alle vragen beantwoord en 60 % goed beantwoord; de zwarte hokjes staan voor de fout beantwoorde vragen. We zeggen dat deze student een stofbeheersing heeft van 60 %. Wanneer deze student een toets moet afleggen over dezelfde stof, weet hij of zij niet welke vragen in de toets worden opgenomen: voor de student ziet iedere toets eruit alsof de toets willekeurig is getrokken uit de afgebeelde verzameling. De willekeur uit zich dan ook in de score of het cijfer voor de toets: al naar gelang de vragen in de toets uitvallen, kan een student die in werkelijkheid 60 % van de stof beheerst de

volgende verschillende resultaten boeken.



Figuur 4. Aantal goed op verschillende toetsen van telkens 50 vragen uit dezelfde vragenverzameling getrokken, bij 60 % stofbeheersing.

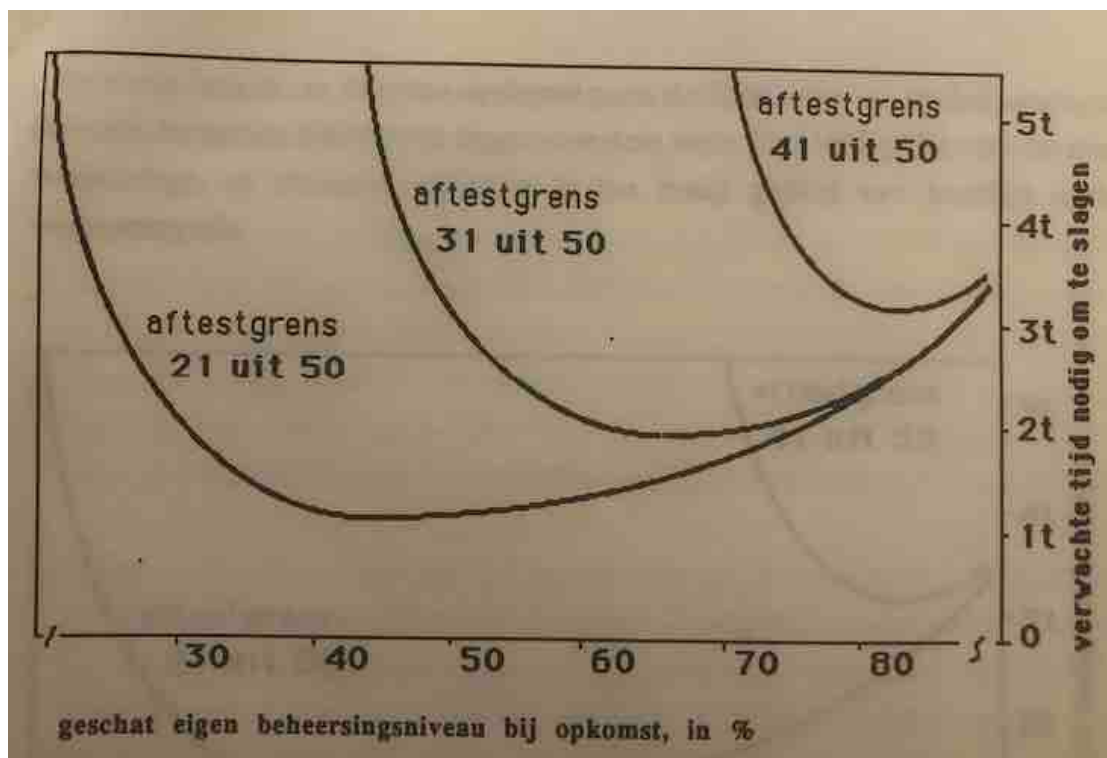
Zoals bij iedere steekproef, kan de student pech of geluk hebben met de toetsvragen zoals hij of zij die krijgt voorgelegd. Rond de aftestgrens maakt een enkel puntje verschil uit of men zakt of slaagt, terwijl het ieder weldenkend mens duidelijk is dat het om volstrekt onbetrouwbare meetverschillen gaat. Differentiële beslissingen rond de aftestgrens kunnen dus niet **inhoudelijk** worden verdedigd, maar wel procedureel: de 'afsprake' is dat voor een voldoende resultaat een exact gespecificeerd aantal punten moet worden gescoord. Deze procedurele interpretatie van de aftestgrens opent de mogelijkheid om door verschuiven van de aftestgrens het gedrag van de studenten te beïnvloeden; dat lijkt een schandelijke gedachte, maar het is de betrokkenen uit te leggen dat deze handelwijze past in het streven naar kwalitatief beter onderwijs dat zeker ook in het belang is van degenen die het onderwijs consumeren.

Merk op dat de **doorzichtigheid** van de toetsing is hier van groot belang is: de student moet in staat zijn zelf het risico te dragen op een enkel puntje verschil toch nog te 'zakken'.

De student heeft in dit model van doorzichtige toetsing tot op zekere hoogte de eigen studieresultaten in de hand, en zal geneigd zijn een 'optimale' studiestrategie te volgen. Wat 'optimaal' is voor de student wordt bepaald door tal van factoren binnen en buiten het onderwijs; men mag veronderstellen dat studenten slechts onder bijzondere voorwaarden bereid zijn meer tijd in de studie te investeren dan nodig is om een normale studievoortgang en dus normale studieresultaten te waarborgen.

De rationele student streeft een niveau van stofbeheersing na dat met de geringste kosten leidt tot slagen. Herkansens kost tijd en energie, maar dat geldt ook voor het nastreven van extra hoge stofbeheersing. Ergens is er een evenwicht tussen beide, daar waar extra studie-inspanning niet meer opweegt tegen de verhoging van de slaagkans die er mee wordt bereikt.

De nu gegeven ingrediënten, doorzichtigheid, de mogelijkheid met aftestgrenzen te schuiven, en naar optimale studiestrategieën tenderende studenten, maken het mogelijk een terugkoppelingsmechanisme te beschrijven voor het beoordelingsproces. Voor de situatie waarin er meerdere herkansingsgelegenheden worden geboden, is een wiskundig model op te stellen op grond waarvan kan worden berekend hoeveel tijd naar verwachting nodig is om voor dat examenonderdeel te slagen.



Figuur 5. Verband tussen stofbeheersing en tijd om te slagen. (B. Wilbrink. Studiestrategieën. Amsterdam: Stichting Centrum voor Onderwijsonderzoek, 1978)

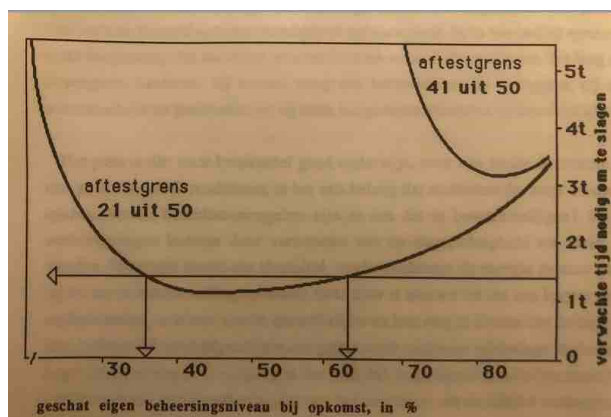
Iedere curve heeft betrekking op een toets van 50 vragen met de daarbij aangegeven aftestgrens. Op de verticale as is de tijd afgezet die nodig is om te slagen, eventueel na herkansen. Op de horizontale as staat de stofbeheersing die de student bij opkomst voor de toets denkt te hebben. U ziet aan het linker deel van iedere curve dat bij lagere stofbeheersing steeds meer tijd nodig is om te slagen, tijd die voornamelijk gaat zitten in het voorbereiden en afleggen van een steeds groter aantal herkansingen. Het rechter deel van de curve laat zien dat ook voor steeds hogere stofbeheersing steeds meer tijd nodig is, maar van deze tijd gaat weinig of niets verloren gaat met herkansen.

De meest voor de hand liggende manier waarop het studiegedrag van de student kan worden beïnvloed, is door het verhogen of verlagen van de aftestgrens, dat zal u niet verbazen.

Er is nu een eenvoudig terugkoppelingsmechanisme: de student stelt zijn studiestrategie af op de eisen die worden gesteld, de docent stelt de eisen dusdanig dat de studenten wenselijke strategieën kiezen. Met studiestrategie wordt hier bedoeld het niveau van stofbeheersing waar de student in de voorbereiding op de toets naar streeft, zoals afgezet op de horizontale as van de figuur. Laten we eens wat nauwkeuriger nagaan wat een wenselijke studiestrategie zou zijn vanuit het gezichtspunt van u als docent.

Er is een laagste, en daarmee optimaal punt, dat is de mate van stofbeheersing die in de kortst mogelijke tijd leidt tot slagen voor deze toets.

Veel belangrijker dan dat formele, wiskundige, te abstracte optimum is het brede gebied van tamelijk optimale studiestrategieën.



Figuur 6. Hoe een zelfde tijdinvestering tot dramatische beheersingsverschillen leidt. (B. Wilbrink. *Beleid bij tentamens en examens*. In T. Vroeijenstijn (redactie) *Kwaliteitsverbetering Hoger Onderwijs*. Voorburg: Stichting Nationaal Congres, 1981).

Bekijk de figuur vanuit de verticale dimensie: de tijd nodig om te slagen. Gegeven een bepaalde benodigde tijd, zijn er twee manieren om in die tijd voor het vak te slagen: trek een horizontale lijn vanuit de gekozen benodigde tijd, en laat vanuit de twee snijpunten van deze lijn met de curve loodlijnen neer op de horizontale as; er zijn twee kiesbare streefniveaus die in de gekozen tijd leiden tot slagen, resp. een (relatief) laag streefniveau, en een (relatief) hoog streefniveau. Bij het lage streefniveau gaat er extra tijd verloren met naar verwachting af te leggen herkansingen, bij het hoge streefniveau wordt er extra tijd besteed aan het bereiken van een hoger stofbeheersing waar als beloning een geringer risico om te zakken tegenover staat. Noem beide strategieën resp. laag-optimaal en hoog-optimaal.

Bij de hoog optimale strategieën wordt een voldoende resultaat in de meeste gevallen zonder herkansen gescoord. Bij de laag optimale strategieën wordt de tijd mede gebruikt voor het afleggen van herkansingen, en is ondanks dat het resultaat nog altijd een lage stofbeheersing. Met andere woorden: studenten die een hoge strategie kiezen, krijgen bij dezelfde tijdinvestering een veel hogere stofbeheersing dan studenten die de lage strategie kiezen, of doortijdgebrek moeten kiezen. Nu kunt u zeggen dat dit voor een enkele toets niet zo geweldig relevant is, maar vergeet niet dat het effect cumulatief werkt: de achterstand of voorsprong wordt bij volgende toetsen vergroot, en wel des te sterker naarmate de leerstof zelf een cumulatieve opbouw heeft. Er is bovendien sprake van een extra inspanning die docenten moeten leveren voor al die studenten die laag optimale strategieën hanteren: zij komen terug om herkansingen af te leggen, zij scheppen roostertechnische problemen, en zij laten jaargroepen studenten op den duur uiteenvallen.

Het punt is dit: voor kwalitatief goed onderwijs, voor een snelle doorstroming, voor een goed numeriek rendement, is het van belang dat

studenten de 'hoge' studiestrategie kiezen. Welke beleidsmaatregelen zijn er om dat te bewerkstelligen? Wel, kleine verschuivingen kunnen door verbeteren van de doorzichtigheid tot stand gebracht worden. Motivatie speelt een sleutelrol, omdat studenten de energie moeten hebben om bij de eerste voorbereiding op iedere toets door te stomen tot dat iets hogere niveau van stofbeheersing, in plaats van de gemakkelijke en luie weg te kiezen van de lage strategie. Een Individueel StudieSysteem is een geforceerde methode: zij dwingt studenten tot een hoge studiestrategie. Overigens is het ook bij Individuele StudieSsystemen zaak te voorkomen dat er tijd verloren gaat met het overdoen van de talrijke tussentoetsen. Een uitstekende methode om verliesgevende lage strategieën uit te bannen is het overgaan op examenregelingen zoals het Amerikaanse Grade Point Average systeem, waar lage en hoge cijfers elkaar vergaand kunnen compenseren en de exameneis wordt geformuleerd als te behalen cijfergemiddelde. Ik geef toe, in het Nederlandse stelsel zitten we nogal vastgebakken aan die rare eis dat in beginsel ieder vak tenminste 'voldoende' moet worden gemaakt, maar ik wijs u erop dat vasthouden aan die eis uw instelling veel geld kan gaan kosten.

Er zijn natuurlijk studenten die binnen de beschikbare tijd niet anders kunnen dan een lage strategie volgen. Dat betekent dat wat zwakkere studenten door een strak opgezette onderwijsprogrammering nog meer op achterstand worden gezet omdat zij naast de normale belasting ook nog de last van het relatief vaak herkansen te verwerken krijgen. Dat wijst in de richting van specifieke programmadiifferentiatie, waarin ook de wat zwakkere broeders en zusters een verstandige studiestrategie kunnen volgen, daardoor langer over de studie gaan doen, maar toch sneller en op een hoger niveau zullen afstuderen dan onder het traditionele rigide stelsel.

Het model maakt het mogelijk om het effect van verhogen of verlagen van de aftestgrens te kwantificeren. Het schuiven met de aftestgrens heeft weinig effect voor degenen met een hoge strategie, maar voor de laag-optimale strategieën leidt het tot drastische effecten. Dat doet vermoeden dat het verstandig is lage aftestgrenzen te vermijden, zoals bij ISS inderdaad gebeurt. •Maar er is geen ISS nodig om eenzelfde gunstige effect te bereiken: uit een verstandig opgebouwde toetsvragenverzameling, die gaat over wat tot de kern van de stof behoort en wat in het onderwijs grondig is behandeld, kunnen toetsen worden getrokken waar redelijkerwijs een hoge beheersing mag worden geeist.

Door de doorzichtigheid te veranderen, zullen studenten hun studiestrategieën aanpassen; dat is hetzelfde mechanisme als het schuiven met de aftestgrens, en kan in beginsel op dezelfde manier worden doorgerekend. Om dat intuïtief duidelijk te maken kunt u als voorbeeld een heel korte toets over een heel moeilijke cursus nemen: zo'n toets is onbetrouwbaar, dat wil zeggen dat een grote studie-inspanning met een beetje pech toch kan leiden tot een heel slechte score, en een gebrekkige voorbereiding door geluk kan worden beloond. Voor zo'n toets zal geen enkele student de moeizame weg volgen, iedereen zal de roulettestrategie gebruiken.

Neem maatregelen die leiden tot meer gewenste studiestrategieën.: kwalitatief verantwoorde toetsvragen, behoorlijke representativiteit van de toets (liever meer dan minder vragen in de toets opnemen), goede oefengelegenheid (proeftoetsen), tevoren bekend gemaakte aftestgrenzen, evenwichtige opbouw van de toets over alle onderdelen van de opgegeven stof, gelegenheid om onmiddellijk te herkansen in plaats van pas na drie maanden, meer mogelijkheden bieden om cijfers elkaar te laten compenseren, en ga zo maar door.

kwaliteit van de toets: enkele praktische mogelijkheden

In mijn Toetsvragen schrijven (Aula 809, online beschikbaar, bw) heb ik een systematiek gegeven voor het construeren van toetsen en het bedenken van toetsvragen. Mijn overweging om een dergelijk boek te willen schrijven is dat de kwaliteit van het hele beoordelingsgebeuren zoals ik dat zojuist heb beschreven, staat of valt met de kwaliteit van de vragen die in de toetsen en examens worden gesteld.

Uit de praktische aanbevelingen in dat boek wil ik de volgende aan u meegeven.

* **Doelstellingen.** Formuleer geen expliciete detaildoelstellingen, maar ga direct over tot het genereren van toetsvragen. De aan te leggen toetsvragenverzameling legt tevens vast wat studenten over de voorgeschreven stof verondersteld worden te weten.

* Die toetsvragenverzameling of **itembank** speelt overigens een belangrijke rol bij doelmatig toetsen: voor het vastleggen van onderwijsdoelen, als oefenmateriaal voor studenten, als mogelijkheid om individuele toetsen te construeren, om voor snel opeenvolgende gelegenheden toetsen te trekken, of als mogelijkheid om items uit te wisselen met andere docenten en andere onderwijsinstellingen.

Bij de automatisering van toetsconstructie en toetsopmaak maakt de itembank het mogelijk zeer snel toetsen samen te stellen, en die in hoogwaardige schriftelijke vorm gereed te hebben voor vermenigvuldiging.

De itembank kan in combinatie met bepaalde constructieregels voor varianten op gegeven toetsvragen, de mogelijkheid bieden tot op zekere hoogte bij iedere nieuwe toets daarin ook een groot aantal vragen op te nemen die in deze bepaalde vorm niet eerder zijn geformuleerd of aan studenten bekend zijn. (de techniek van rompvragen)

* **constructieregels** voor het genereren van toetsvragen. Een gigantisch struikelblok voor veel docenten is het corvee telkens weer nieuwe vragen over dezelfde stof te moeten bedenken. Dat gaat ook wel eens behoorlijk fout, doordat in de loop der tijd steeds meer vragen worden bedacht die de kern van de stof niet meer raken, die uitsluitend over marginale details gaan. De docent moet de beschikking hebben over regels aan de hand waarvan nieuwe vragen over dezelfde leerstof min of meer kunnen worden geconstrueerd, op ambachtelijke wijze uitgeschreven zou je kunnen

zeggen. Dergelijke constructieregels heb ik in mijn boekje 'Toetsvragen schrijven' proberen te geven. Het zijn natuurlijk geen blind toepasbare regels: er wordt wel gevraagd dat u een grondige leerstofanalyse maakt

Doorzichtige toetsing, heldere constructieregels, leiden als vanzelf tot een uitgesproken didactiek voor de betreffende leerstof, een didactiek die niet zozeer berust op psychologie, alswel op wetenschapsfilosofie: uit welke begrippen is de leerstof opgebouwd, welke relaties of wetmatigheden tussen deze begrippen zijn van belang, hoe zijn meerdere relaties of wetmatigheden in een theoretisch netwerk verbonden, welk soort definities zijn voor welke begrippen zinvol te gebruiken. Vul dat aan met het beginsel dat onderwijs nimmer alleen abstracties behoort aan te bieden, maar altijd tenminste de abstracte termen voor de studenten hanteerbaar moet maken door talrijke voorbeelden te geven en te vragen. Het ben en van het concrete leidt tot een hogere motivatie en betrokkenheid van de studenten, maakt het bovendien mogelijk op veel doelmatiger wijze toetsvragen te construeren.

* **keuzetoetsen** in het bijzonder: inderdaad, vierkeuzevragen zijn vaak ondoelmatig omdat van docenten niet verwacht kan worden dat zij naast het juiste alternatief ook nog meerdere onjuiste alternatieven bedenken die technisch en didactisch aanvaardbaar zijn. Het is daarom veel doelmatiger, vasthoudend aan de keuzevraagvorm, om met drie of zelfs twee alternatieven te werken. U kunt dan toch gebruik blijven maken van computerprogrammatuur die op vierkeuzevragen is afgestemd omdat die computer als dom werktuig het verschil niet zal merken. De vierkeuzevraag is een kunstgreep geweest om in de zestiger jaren het CITO een vliegende start te geven, maar is zeker geen norm voor hoe een goede keuzevraag eruit moet zien.

https://benwilbrink.nl/publicaties/87efficient_toetsen.pdf