

rapport 11-71 48 blz. offset Ben Wilbrink: Levert selectie in het hoger onderwijs iets op?

[Notitie gemaakt voor RWO-bijeenkomst Helvoirt, als ik mij goed herinner] Ad hoc inventarisatie van de belangrijkste overwegingen, van belang voor het evalueren van de mogelijkheid en wenselijkheid van selectie in het hoger onderwijs via propedeutische examens.

In verband met de ingewikkeldheid van de problematiek en tekort aan voorbereidingstijd heeft de tijd voor een behoorlijke eindredactie ontbroken, vandaar stijl en type-fouten en soms minder goed geformuleerde argumentaties. Aan een meer definitieve versie zal de komende maanden gewerkt worden.

3

INLEIDING,

Dit rapport gaat over selectie van (aankomende) studenten. Deze groep is zélf al het produkt van een langdurig proces van selectie tijdens lager en middelbaar onderwijs. Deze selectievooraf draagt vele kenmerken van sociale selectie. Daarom ook is de opmerking dat het grootste deel van de Nederlanders van 18 jaar en even ouder in het geheel niet voor universitaire selectie in aanmerking komt, een schrijnende opmerking.

Ook het engelse Crowther Report (1959)

signaleerde dit: "The report is about the education of English boys and girls aged from 15 to 18. Most of them are not being educated". Wie over selectie in of voor het hoger onderwijs spreekt, moet ook spreken over de nota's van Posthumus en het wetsvoorstel herstructurering. Een deel van het betoog zal dan ook gepresenteerd worden als reactie op geciteerde uitspraken uit genoemde stukken. Een poging tot evaluatie van het wetsvoorstel, voorzover dit betrekking heeft op selectieve momenten in het onderwijs, zal gedaan worden. Daarbij zullen kosten en baten van huidige selectie-methoden (die vrijwel gelijk zijn aan de selectie zoals die uit het wetsvoorstel volgt) tegen elkaar afgewogen worden. Deze afweging zal voorlopig kwalitatief moeten zijn, omdat vele onderwijsresultaten moeilijk kwantificeerbaar blijken.

Het selectieprobleem zal benaderd worden vanuit de inzichten die de psychometrie biedt ten aanzien van de technische problemen die zich voordoen, en vanuit onderwijskundige standpunten waar het er immers om gaat de selectie in te schakelen ten gunste van "het onderwijs", en waar het hanteren van selectieve beoordelingsmethoden in het onderwijs zeker van (nadelige) invloed is op de kwaliteit van dat onderwijs.

Wanneer men nederlands onderzoek over selectie of studieduur in het hoger onderwijs raadpleegt, blijkt verrassenderwijs zowel een behoorlijke psychometrische als een onderwijskundige analyse veelal te ontbreken. Dit heeft tot gevolg dat een aanzienlijk deel van hier verricht onderzoek resultaten heeft opgeleverd die een zeer beperkte waarde hebben. Vaak is ook sprake van een overdreven operationalisme, veel verslaggevingen geven de indruk niet over mensen te handelen, maar over cijfers en correlaties die van al hun menselijke betekenis ontdaan zijn.

Aangezien de selectie-omstandigheden in het nederlandse hoger onderwijs nogal verschillen van die in andere landen, is het moeilijk het komende betoog met empirisch materiaal te ondersteunen, omdat ook nederlandse onderzoek gegevens om bovengenoemde redenen meest irrelevant blijken. Dat betekent dat het betoog meer zal steunen op theoretische verworvenheden, wat geen bezwaar lijkt omdat uit de theorie en enkele redelijke aannamen over onze onderwijs situatie toch vrij harde conclusies te trekken zijn.

De vraag naar de mogelijkheden of onmogelijkheden van selectie kan alleen op zinvolle wijze onderzocht worden wanneer we

weten waarom we selekteren. De memorie van toelichting op het ontwerp van wet noemt drie redenen dat "zowel een periodieke beoordeling als een beperking van de verblijfsduur noodzakelijk zijn".

(1) *"samenwerking binnen een groep is alleen mogelijk wanneer tussen kundigheden, kennis, toewijding en toekomstplannen niet te grote onderlinge verschillen bestaan"*.

Commentaar: In zijn algemeenheid is dit geen onjuiste uitspraak, de vraag is echter op **welk moment welke verschillen** te groot worden om in beaalde mate samen te werken. Voor het lager onderwijs bijv. is het niet zo dat het gehanteerde systeem van zittenblijven resulteert in kleinere onderlinge verschillen tussen leerlingen in dezelfde klas, zie onderzoek vermeld door Cook (1951) en Doornbos (1970).

Ook is in het onderwijs een grote verscheidenheid vaak een gewenst goed, en vooral in het wetenschappelijk onderwijs waar vaak door argument en tegen-argument inzicht in een wetenschapsgebied verworven moet worden. Ook gaat dit argument uit de memorie van toelichting voorbij aan de vraag of binnen het onderwijs al het mogelijke gedaan is om ondanks altijd bestaande verschillen een goede samenwerking te realiseren.

Tenslotte lijkt hier niet zozeer de samenwerking van studenten onderling bedoeld te zijn, ons onderwijs werkt immers nog steeds in hoge mate individualiserend, meer eerder de "samenwerking"

5

die bestaat uit de confrontatie tussen de docent en een groep studenten; m.a.w. blijkt hier voor te liggen een opvatting van onderwijs waarin de docent de actieve, en de student samen met zijn medestudenten een passieve rol heeft.

(2) *"een maatschappij, die op een verdeling van arbeid berust, verwacht terecht dat diegenen die jonge mensen gedurende vele jaren van hun studie hebben begeleid en gadegeslagen een oordeel uitspreken over die kundigheden, kennis en toewijding en dat zij daarbij redelijke normen als toetsingscriteria voor een afsluitend examen hebben gesteld"*.

Commentaar: Moeten we hieruit opmaken dat de maatschappij "terecht verwacht" dat er in het onderwijs geselecteerd wordt? Daar ben ik nog niet zo zeker van. Wél is zeker dat juist tegen dit soort ideologie steeds grotere bezwaren bestaan juist bij dát deel van "de samenleving" dat onderwijs volgt. Maar afgezien van de twijfelachtige ideologie achter dit argument moeten er twee belangrijke vragen over deze uitspraak gesteld worden Ten eerste is het nog

maar de vraag of de docenten een gefundeerd oordeel uit kunnen spreken, uit de onderzoeken die zich met dergelijke oordeelsvormingen hebben beziggehouden blijkt dit zeker niet evident te zijn (zie bijv. Cronbach, 1970). Ten tweede ligt het veeleer voor de hand de bedoelde garantie van bekwaamheid te zoeken in de kwaliteit van het onderwijs, dan in veelal weinig zeggende beoordelingen en examenresultaten. M.a.w., van een onderwijsinstelling verwacht men toch dat er in de eerste plaats onderwijs gegeven wordt, niet dat er in de eerste plaats geselecteerd wordt?

(3) "de snelle en versnelde "schaalvergroting" van het gehele onderwijs maakt het noodzakelijk dat grenzen worden gesteld aan het in volle dagtaak, zonder zorg voor dagelijks levensonderhoud, deelnemen aan dat onderwijs"

Commentaar: Mogen we hieruit concluderen dat er geselecteerd moet worden omdat de aantallen beperkt moeten worden? Dat zou in ieder geval het enige juiste en eerlijke argument zijn, tenminste zolang redeneringen die zich baseren op "geschiktheid" van studenten niet aantoonbaar juist zijn.

6

Mede op grond van het voorgaande zullen we er

voorlopig van uitgaan dat selectie als doel heeft het beperken van het aantal studenten. Het geschiktheids-argument vatten we niet op als een argument om te selekteren zonder meer, maar als een aangeven van de wijze waarop we die selectie willen uitvoeren: iedereen wil graag de meer "geschikten" aannemen en de overigen afwijzen.

We gaan er eveneens voorlopig van uit dat zoiets als "geschiktheid" voor de studie inderdaad bestaat, dat het in de uitslag van een propedeutisch examen "blijkt" en dat het de enige relevante variabele is op grond waarvan we een selectie kunnen en willen uitvoeren die betere resultaten oplevert dan bijvoorbeeld een lotingsprocedure. Of een en ander houdbaar is vormt onderwerp van deze beschouwing.

Posthumus (2e voortgangsnota blz.37).

"De eerste taak is dan: het opstellen van definities en het uitwerken van meetmethoden voor de drie intensieve grootheden: peil van het onderwijs, gemiddelde studiegeschiktheid der studenten, moeilijkheid van de examens.

Ik kom nog niet verder dan het antwoord, dat ik zo vaak op een tentamen heb gekregen: 'ik weet het wel, maar ik kan het niet zeggen'. Wij moeten voorlopig opereren met de best beschikbare definitie. Wij meten zowel de

geschiktheid voor de studie als het peil van het onderwijs aan de uitslagen der examens. De instelling is geneigd hoge examenuitkomsten als resultaten van het onderwijs te beschouwen. Studenten zien daarin graag een aanwijzing voor hun geschiktheid. De buitenstaander daarentegen neigt er toe de uitslagen te beschouwen als maatstaf voor de moeilijkheid der examens".

In dit rapport zal geprobeerd worden te formuleren wat Posthumus niet kan zeggen. Of we inderdaad kunnen opereren met de "best beschikbare definitie" is ook een vraag waarop een antwoord gezocht zal worden, dit antwoord zal in belangrijke mate afhangen van de mogelijkheid om meetmethoden voor "peil", "geschiktheid" en "moeilijkheid" te vinden. Wat uit de literatuur over de aard en de waarde van dergelijke meetmethoden bekend is, zal worden geïventariseerd.

7

WIE SELEKTEREN?

De memorie van toelichting (9-1):

"wanneer wij spreken over selectie, denken wij geenszins aan een eenzijdige handeling vanwege de docent. In belangrijke mate zal de student zelf door een betere ordening en door verbeterde representativiteit meer dan vroeger in staat zijn zich te "selecteren". Hij zal meer en

doorgaans eerder inzicht krijgen in eigen aanleg en motivatie. Ook in sociaal opzicht vormt de voorgestelde regeling, die beoogt van de propedeuse een oriëntatie fase te maken in de ware zin van het woord, dan ook een verbetering die te gauw wordt onderschat"

Het propedeutisch examen is het belangrijkste instrument daarbij.

De memorie van toelichting (3-8):

"Het propedeutisch examen vervult - als elk examen - een selecterende functies zowel de student als de faculteit dienen zich zo spoedig mogelijk een oordeel te vormen over de geschiktheid en de belangstelling voor de gekozen studierichting, zodat zowel zelf-selectie als consilium abeundi mogelijk zijn".

Het hanteren van de term als "zelf-selectie" is erg makkelijk, maar we moeten ons afvragen hoe een dergelijke "zelf-selectie" verloopt, welke momenten daarbij bepalend zijn. Daarbij zal vooral het verband tassen de vormgeving aan het onderwijs en de wijze waarop zelf - selecterende beslissingen kunnen worden genomen, moeten worden nagegaan.

Een tweede belangrijke vraag, die we daarna zullen bespreken, is de verenigbaarheid van zelf-selectie en selectie vanwege de onderwijsinstelling. Het vermoeden bestaat namelijk dat selectie door de docent een reële

vorm van zelf selectie door de student verhindert. Dit "vermoeden" zullen we meer concreet moeten maken.

Het eerste wat opgemerkt moet worden over zelf selectie, is dat ook de student die "zichzelf selekteert" de beschikking moet hebben over alle informatie die relevant is voor zijn beslissing.

Dit is een opmerking die niet zal verrassen, hoewel de implicaties ervan nogal vérstrekkend zijn. Immers, we moeten ons bij de zelf selectie evenals bij de selectie door de onderwijsinstelling, afvragen of de daarbij gebruikte informatie inderdaad relevant, "geldig", is voor de genomen beslissing. Het grootste deel van dit rapport is besteed aan de vraag naar de "geldigheid" van de door de onderwijsinstelling gepleegde selectie, ook voor de zelf-selectie zullen we in principe alle daar besproken problemen moeten beschouwen. We zullen dat hier niet doen, en volstaan met er op te wijzen dat de mate van "geldigheid" die verkregen kan worden in de selectie in zekere zin ook een bovenste grens vormt voor de mate van geldigheid van de zelf-selectie procedure. Met enige voorzichtigheid is al hetgeen volgt over selectie door de onderwijsinstelling eveneens van toepassing op selectie van de student op zichzelf. Wanneer de vooruitzichten op

effektieve selectie procedures slecht blijken te zijn, mogen we ook niet veel verwachten over de mate waarin zelf-selectie zinvol kan zijn.

Het bovenstaande is een belangrijke overweging die we voortdurend voor ogen moeten houden, hoewel er in het volgende niet meer op terug gekomen zal worden.

De zinvolheid van zelf-selectie en van selectie brengen ons inziens dus zeer nauw samen, omdat van goeddeels dezelfde informatie gebruik gemaakt moet worden, en de beslissing van gelijke aard is. Helemaal gelijk is de situatie echter niet, zoals blijkt in het hoofdstukje "Kritiek op het gehanteerde selectiemodel". Wanneer we de student in zijn eerste studiejaar optimaal in de gelegenheid willen stellen om op goede gronden voor zichzelf te beslissen of hij deze studie zal voortzetten of niet, moet het onderwijsprogramma daar ook op afgestemd worden. In dat verband wordt dan gesproken van "representativiteit", en "oriënterend karakter" van de propedeuse. Laten we ons beperken tot de representativiteit, en dit begrip proberen te concretiseren. Het gaat er om dat de student zich een goede indruk kan vormen van het vak, de studie, en de eisen die vak en studie aan de student stellen.

Zodoende kan hij nagaan of zijn verwachtingen

overeenstemmen met de "realiteit", en wanneer er geen te groot verschil is, zal hij besluiten door de studeren. Nu is zowel het verrichten van wetenschappelijk onderzoek als het verrichten van toegepast wetenschappelijk werk een dermate complexe activiteit, met bovendien een moeilijk invoelbare belevingsdimensie, dat iedere hoop om iets daarvan in een propedeuse te vangen ijdel zal blijken.

Dan blijven over de studie en de eisen. Is het mogelijk in een propedeutisch jaar, waarin de aankomende studenten met honderden tegelijk door een zeer schools college en tentamenprogramma heen gehaald worden, is het mogelijk in een dergelijke situatie een goede indruk te geven over wat het doctoraalprogramma aan mogelijkheden biedt? Vele docenten zullen dat moeten betwijfelen. Resteren de eisen, de vraag of ik de studie zal kunnen "bolwerken". De nota's van Posthumus evenals de memorie van toelichting wemelen van de uitspraken over de relativiteit en de verschuifbaarheid van "normen". Als dit inderdaad de realiteit is zal de student zich moeten afvragen hoe hij uit de competitie-slag met zijn medestudenten te voorschijn zal komen, en óf hij bepaalde risico's om te "mislukken" wel wil aanvaarden. Welnu, het is duidelijk dat dergelijke overwegingen geen

overtuigende indruk van de rationaliteit van self-selektie kunnen geven.

Het spijt me dat het voorgaande zo negatief uitvalt, maar we zullen weinig anders kunnen doen dan constateren dat iedere beslissing die de student neemt over zijn al dan niet verder studeren gebaseerd is op onvoldoende informatie, een situatie waarin wel enige verbetering gebracht zou kunnen worden, maar die toch in welke vorm die we aan ons hoger onderwijs ook geven, tot op grote hoogte onvermijdelijk is. En waar ons onderwijs in wezen competitief is ingesteld zullen overwegingen t.a.v. "slaagkansen" een overheersende rol gaan spelen, zeer ten nadele van de betrokkenen én van het onderwijs. Dat namelijk deze slaagkansen moeilijk voorspelbaar zijn zal blijken uit dit rapport.

Het voorgaande staat wel in schrille tegenstelling tot de uitspraak van Posthumus (1e voortgangsnota, blz 38):

"In het voorgaande hoofdstuk over 'het afnemen van examens heb ik betoogd dat selectie moet betekenen: de optimale informatie over iemands mogelijkheden, zodat hij voor

- 10-

zichzelf en daarmee hopelijk ook voor de gemeenschap een zo verantwoord mogelijke keuze kan doen. Selectie moet met andere

woorden de optimale voorwaarde voor zelfrealisatie scheppen".

In tegenstelling hiermee komen wij tot de conclusie dat zelf-selektie plaatsvindt op basis van onvolledige en irrelevante informatie, waarbij weliswaar de onvolledigheid geen obstakel hoeft te vormen, maar wel de irrelevantie van overwegingen t.a.v. slaagkansen, gezien de wijze waarop in ons huidige onderwijs deze slaagkansen bepaald worden (zie daarvoor o.a. "Rapport van de commissie studieduur" van de Academische Raad, 1964, het verslag van de zgn. "commissie-Posthumus").

Het tweede punt: zijn zelf-selektie en selektie met elkaar te verenigen? Ze worden door velen gezien als samenhangend, bijv. Posthumus (de universiteit etc. blz. 57):

"De functie van het propedeutisch onderwijs en het propedeutisch examen wordt als volgt gezien:
<ol class='a'>a. het verschaffen van inzicht aan de student in de methoden en de doelstellingen van de door hem gekozen studierichting; b. de vorming van een oordeel door de student en door de faculteit omtrent de geschiktheid voor de gekozen studierichting. Getracht moet worden propedeutisch onderwijs- en examen-programma's op te stellen, die beantwoorden

aan voorwaarden van representativiteit en selectiviteit".

De manier waarop Posthumus de selectie zoals die uiteindelijk plaatsvindt omschrijft, doet al enige tegenstrijdigheid vermoeden (De universiteit etc., blz. 14:)

"Voor vrijwel alle examens geldt dat de vereiste prestaties door voortgezette studie kunnen worden verbeterd. Voorts kunnen vrijwel alle examens onbeperkt worden herhaald. De afgewezen kandidaat staat dus telkenmale voor de beslissing: herhalen of staken. De som van deze persoonlijke beslissingen bepaalt de statistische uitkomsten van de studieduur der geslaagden en het percentage van de afzwaaiers. Faculteiten en studenten tezamen vervullen daarmee een functie, die vaak "selecterend" wordt genoemd".

Docenten en studenten bepalen gezamenlijk de selectie. Het boze vermoeden rijst dat deze samenwerking hierin bestaat dat de docenten selekteren en de studenten zich laten selekteren.

Wanneer de onderwijsinstelling gaat selekteren na bijvoorbeeld een jaar studie, is het onvermijdelijk dat het studiegedrag van de studenten gericht is op het examen, op een competitieve strijd met hun medestudenten. Vrijwel alle ruimte en gelegenheid tot

oriëntering dreigt hierdoor opgeslokt te worden, een dreiging waartegen nóch het ontwerp noch de memorie van toelichting enige waarborgen schept. Wat in de praktijk zal gebeuren, en ook nu al op grote schaal gebeurt, is dat het propedeutisch jaar volgestopt wordt met programma's gericht op het verwerven van basiskennis, waarover aan het eind van het jaar op selectieve wijze getoetst gaat worden. Waar blijft de oriëntatie, waar vinden we posthumus' "zelf-realisatie", waar blijft de "representativiteit" van het eerste jaar? Aan het eind van het jaar vindt de grote wedstrijd plaats (Posthumus, De universiteit etc., blz.34):

*"Het vergelijkend onderzoek naar de **geschiktheid** vindt plaats door de prestaties van de studenten, die **hetzelfde** onderwijs hebben gevolgd, met dezelfde maatstaven te beoordelen".*

Waar Posthumus bovendien nog voorstelt dat het examen bij voorkeur moet bestaan uit meetkeuze studietoetsen, weten we helemaal zeker dat er van zelf-selectie op basis van een behoorlijke oriëntatie van de student en representativiteit van het onderwijsprogramma, niets meer overblijft.

Concluderend t.a.v. de vraag naar verenigbaarheid van zelf-selectie en selectie door de

onderwijsinstelling, moeten we stellen dat iedere vorm van pressie die op de student uitgeoefend wordt door studieprogramma en inrichting van het examen ten koste gaat van de rationaliteit van zelf-selekerende beslissingen. Willen we de nadruk leggen op zelf-selektie, dan zullen we daar ook alle gelegenheid toe moeten scheppen door selectie van bovenaf achterwege te laten. Is dat niet het geval dan zijn alle zogenaamde zelf-selekerende beslissingen in feite door de onderwijs-instelling afgedwongen, en dus géén zelf -selectie.

12

WAAROP SELEKTEREN?

Zoals gezegd gaan we er van uit, dat de selectie plaatsvindt omdat er niet voor iedereen plaats is. Er moet vervolgens een selectie-procedure gevonden worden, een lotings-procedure of, als we daar meer baat bij denken te hebben, een procedure waarbij we gebruik maken van al bekende of eventueel in te winnen informatie over de aankomende (of eerstejaars) studenten.

Een eerste selectie-drempel, het zij volledigheidshalve vermeld, is een eindexamen dat recht op toelating tot het hoger onderwijs geeft. Dit is overigens geenszins een triviaal gegeven, zoals nog zal blijken. Een deel van

deze gelukkigen zal een plaatsje in een instelling voor hoger onderwijs ambiëren, en voor deze groep is ons selectieprobleem relevant.

Of het de beste procedure is om te selecteren op grond van een propedeutisch examen kunnen we nu nog niet zeggen, we gaan er wel voorlopig van uit dat, in overeenstemming met het ontwerp van wet, de propedeutische als selectie instrument gehanteerd wordt.

Allereerst zullen we daarbij selecteren op "uithoudingsvermogen" in het eerste jaar, waaruit dat dan ook mag bestaan. Het komt veel voor dat studenten wel onderwijs volgen maar niet opkomen voor een tentamen of examen, ofwel omdat zij juist de studie opgegeven hebben, ofwel omdat ze zichzelf weinig kans geven en nog liever even uitstellen. Voor de eerste categorie geldt dat in feite de selectie-beslissing al gevallen is vóór het examen, en we zullen er achter moeten komen op welke gronden mensen menen een studie voortijdig te moeten staken omdat deze redenen veel of alles van doen hebben met de studie-omgeving en het gegeven onderwijs, zaken waarvoor de faculteit een grote verantwoordelijkheid draagt. Het is niet juist gevallen van vroeg afbreken van de studie zonder meer "zelf-selectie" te noemen omdat

invloeden vanuit de onderwijssituatie hiertoe kunnen dwingen. Voor de tweede categorie, de uitstellers, geldt dat zij daarmee ook de selectie-procedure uitstellen zij het dat hierdoor een grotere aanspraak op hun uithoudingsvermogen gedaan wordt, en andere omstandigheden (de beruchte "voor-zieningen pakketten" die ingetrokken kunnen worden) een toenemende

pressie op deze studenten zullen uitoefenen. Hun situatie is dan ook in hoge mate vergelijkbaar met die van de "gezakten" die zich op een herkansing voorbereiden. Het bovenstaande moeten we o.a. dáárom signaleren, omdat hiervan veelal in allerlei statistieken niets terug te vinden is, en het verschijnsel een zeer belangrijk kenmerk kan zijn van de selectie zoals die, soms geheel ongewild, in de praktijk plaatsvindt. Uit de volgende hoofdstukken zal blijken dat zoiets als "geschiktheid voor de studie geen 44nduidig begrip is, en, niet eenvoudig valt te voorspellen. Voor het ogenblik echter gaan we er van uit dat er mee gewerkt kan worden. "Geschiktheid" wordt door velen gezien als de eigenschap die de belangrijkste bepalende factor in het toelatingsbeleid moet zijn. Zo bijv. Posthumus (De universiteit, blz. 34):

"Het vergelijkend onderzoek naar de geschiktheid vindt plaats door de prestaties van studenten, die hetzelfde odderwijs hebben gevolgd, met dezelfde maatstaven te beoordelen".

Laten we nagaan hoe de situatie is wanneer de werkwijze van Posthumus gevolgd wordt.

Allereerst het punt van dezelfde maatstaven. Dit lijkt een een-voudige en rechtvaardige zaak, maar pas op. Heel bekend is het verschijnsel dat op verschillende basis gevormde deelgroepen verschillende resultaten op dezelfde test behalen. Zo zijn er vrij wel altijd systematische verschillen in test-scores tussen mannen en vrouwen. Voor propedeutische examens zijn er systematische verschillen tussen de groepen die op basis van genoten vooropleiding gegroepeerd zijn. Talloze fijnere verdelingen, of verdelingen op basis van persoonlijkheids-, motivatie-, intelligentie-, en andere gegevens zijn te maken waarbij veelal systematische verschillen in test-scores waarneembaar zijn. Wanneer dergelijke systematische verschillen geen invloed op de selectiebeslissingen mogen hebben, zullen we bij een selectieprocedure daarmee rekening moeten houden, en voor diverse onderscheidbare groepen verschillende maatstaven aanleggen. Op grond van deze overwegingen moeten we onderzoeken voor

welke groepen we de slaag-zak grens omhoog of omlaag moeten brengen. Als we dit inderdaad gaan doen zal er ongetwijfeld tegen geprotesteerd gaan worden. Wanneer we het niet doen echter zullen we meer foutieve beslissingen nemen dan anders het geval zou zijn, en zullen we grotere onrechtvaardigheden tegenover de betreffende studenten begaan. De "Standards for Educational and Psychological Tests and Manuals (1966) (voortaan geciteerd als "Standards") beveelt sterk aan om onderzoek te doen naar de validiteit van een test-procedure voor verschillende deelgroepen wanneer het vermoeden bestaat dat er verschillende kunnen bestaan tussen deze deel-groepen (blz.20).

Belangrijker misschien dan de genoemde systematische groepsverschillen zijn individuele verschillen die niet relevant zijn voor de selectie-beslissing, maar die wél beïnvloeden. Hier zullen persoonlijkheids-, motivatie-, en intelligentie-verschillen een grote rol spelen bij de kans die men heeft om afgewezen te worden, terwijl dergelijke verschillen voor het onderwijs meestal niet relevant zullen zijn. (Dit is een bron van onbetrouwbaarheid voor de examen-uitslagen). Om het te illustreren aan de intelligentie: zoals langzamerhand al meer algemeen bekend,

bestaat er niet zoiets als "de" intelligentie, maar kunnen we vele verschillende factoren onderscheiden binnen het verzamel-begrip "intelligentie". Een heel vervelend verschijnsel waar we mee te maken hebben in het onderwijs is dat sterk geselecteerd wordt of factoren als verbale vaardigheden, die bijvoorbeeld in latere beroepspraktijk van veel minder belang kunnen zijn. Iets dergelijks doet zich bij ieder examen in feite voor: er wordt een sterker beroep gedaan op bepaalde intelligentie-factoren ten koste van andere, terwijl in het voorafgaande of het volgende onderwijs deze intelligentie-factoren vaak van minder belang zullen zijn. Een examen zal op deze wijze nooit optimaal kunnen zijn, terwijl bovendien bepaalde individuen systematisch in het nadeel zijn op examens, terwijl dat nadeel in de onderwijs-situatie niet relevant is.

15

Niet alleen de meer standvastige variabelen als de bovengenoemde zijn van invloed, ook allerlei verschillen in voorbereiding(stijd) die al of niet toevallig ontstaan zijn, spelen een storende rol. Als belangrijk voorbeeld kunnen we milieu-factoren noemen, die vaak een sterk verband houden met de kansen om uitgeselecteerd te worden (of zelf-selektie te moeten plegen) terwijl de feitelijke capaciteiten daartoe geen

aanleiding zouden geven. (zie wederom de studie van Doornbos die relevante onderzoeken bespreekt).

Hetgeen tot nu toe gezegd is over de gelijkheid van maatstaven heeft in feite betrekking op de metingen die aan deze maatstaven gerelateerd zijn, in casu het examen. Deze examenresultaten zijn in vaak forse mate onbetrouwbaar, een aantal bronnen van onbetrouwbaarheid hebben we juist genoemd, er zijn nog vele andere factoren (voor een systematische opsomming zie Schwarz (1971), die aan onbetrouwbaarheid van de metingen bijdragen.

In dit rapport zal overigens weinig aandacht aan het vraagstuk van de betrouwbaarheid van toetsen etc. besteed worden, in tegenstelling tot wat in nederlandse literatuur gebruikelijk lijkt. Betrouwbaarheid zegt alleen iets over de vraag of we al dan niet iets "meten", en is als zodanig van belang als ondersteuning voor een analyse en schatting van inhoudelijke, voorspellende, en "construct" geldigheid, maar kan deze nooit vervangen. ("Standards" blz. 28). Hoe groter de onbetrouwbaarheid is, des te minder waarde kunnen we aan de examenresultaten hechten. De "waarde" van deze resultaten wordt o.a. bepaald door de mate waarin studie-scores e.d. "voorspelbaar"

is op basis van deze propedeuse-resultaten (waarover in de volgende hoofdstukken meer). Een andere factor die van belang is voor het bepalen van de "waarde" van de examenprocedure is het onderzoek naar de mate waarin we met onze examenprocedure inderdaad iets als "geschiktheid" meten. (Het probleem van de "construct" geldigheid, zie bijv. de "Standards") we zullen dan ook heel nauwgezet moeten formuleren wat we niet[sic] onder geschiktheid verstaan voordat we een examen opstellen.

16

In de mate waarin we hiermee nalatig zijn, zullen we examen-resultaten krijgen waarvan niet geheel duidelijk is hoe we ze moeten interpreteren, en die interpretatie is nodig omdat we op basis van deze resultaten moeten beslissen iemand af te wijzen of te accepteren. Bovendien zal vrij uitgebreid onderzocht moeten worden het verband tussen gedragingen en prestaties van studenten tijdens de studie en de behaalde examen-resultaten, omdat beide verband met elkaar moeten hebben wil er van een zinvolle examenprocedure sprake zijn. (Dit is dus iets anders dan het nog te bespreken verband tussen examen-resultaten en uiteindelijk studiesucces in de uitoefening van het vak).

Het betreft hier een problematiek die uitgebreid behandeld wordt in literatuur over "Educational Measurement", bijv. het gelijknamige boek verzorgd door Thorndike (1971), of De Groot, Van Naerssen e.d. (1969). Onder verwijzing naar deze literatuur zullen we er hier verder het zwijgen toe doen.

In het wetsvoorstel wordt er stilzwijgend van uitgegaan dat selectie plaatsvindt op dezelfde wijze als nu gebeurt: door middel van tentamens en examens, al of niet in de vorm van meerkeuze-toetsen.

Zoals uit het voorgaande al gebleken is, betekent dit geenszins dat niet tevens op variabelen als vooropleiding, intelligentie, geslacht, milieu van afkomst etc. geselecteerd wordt. Dit zijn namelijk gegevens die verband houden met studieresultaat als gemeten door tentamens (ze correleren er mee). Wanneer we selekteren op examenresultaten, selekteren we in feite ook op alle variabelen die met deze examenresultaten verband houden. (Uitgebreid besproken in Lord & Novick 1968). Omdat het zeer ongewenst is persoonlijkheids, biografische e.d. gegevens te gebruiken als gegevens voor selectie, moeten we onderzoeken in hoeverre examenresultaten hierdoor bepaald worden om dit mede te betrekken in studie van de vraag of een

dergelijke selectie d.m.v. examens wel wenselijk is. De memorie van toelichting is op dit punt van een gevaarlijke vaagheid, waartegen scherp geprotesteerd moet worden (1-5):

"Bij deze selectie kunnen vele kenmerken, afzonderlijk of tezamen, worden gehanteerd: Kalenderleeftijd, sekse, lichamelijke of geestelijke eigenschappen, maatschappelijk milieu, kundigheden, vorderingen, diplomabezit, toekomstplannen".

Wanneer we gebruik maken van tentamenresultaten voor de selectie, houden we rekening met de mate waarin de student zich leerstof in het eerste jaar voldoende heeft weten eigen te maken. Dit is een merkwaardige constructie, omdat het eerste studiejaar nogal verschilt van de rest van de studie. Het studieprogramma is voor iedereen gelijk, terwijl voor het doctoraal programma een zo groot mogelijke keuze-vrijheid voorgesteld wordt. Dit is op z'n minst zeer inkonsekvent, maar betekent bovendien dat we voorzichtig moeten zijn met zonder meer te veronderstellen, dat examenresultaten over een gedwongen uniform studiepakket iets zeggen over te verwachten studiegedrag wanneer we de student vrijlaten. Een andere belangrijke overweging is dat de schoolsheid van het

eerste studiejaar in schril contrast staat tot de pretenties van wetenschappelijk onderwijs t.a.v. kritisch, zelfstandig, creatief etc. werk tijdens zowel als na de studie. Wanneer eerst geselecteerd wordt op schoolsheid kun je niet verwachten daarmee tevens een creatieve kritische etc. groep studenten aangenomen te hebben.

Dit zijn vrij ernstige tekortkomingen, zowel in de wijze waarop selectie vandaag de dag plaatsvindt, als waarop deze volgens Posthumus zal verlopen. Op deze wijze bereiken we een ongetwijfeld ongewenste uniformering van de studenten-populatie. Zeker voor het wetenschappelijk onderwijs is het daarentegen zeer gewenst een groot scala van interesses, capaciteiten etc. bij de studenten te bevorderen. Dit effect van ongewenste homogenisering van groepen is een bezwaar dat in het algemeen tegen massale, op tests of toetsen gebaseerde selectie gemaakt kan worden, omdat het een groot stuk onrechtvaardigheid t.a.v. individuele personen incorporeert. Bij selectie voor het hoger onderwijs komt daar nog bij dat ook aan de maatschappij op deze wijze schade toegebracht kan worden. Ook Hazewinkel (1971) waarschuwt voor dit gevaar.

Als afsluiting van dit hoofdstuk signaleren we nog

een belangrijke tekortkoming van het gebruik van tentamenresultaten als selectie gegevens. De student bereidt zich op een tentamen veelal "kort en hevig" voor, hij is in het ideale geval tijdens de tentamenafname op het toppunt van zijn kunnen. Indien vrij korte tijd later (een week, een paar maanden) onverwachts hetzelfde tentamen nogmaals

- 18 -

afgenomen wordt, zullen we zien dat zijn prestaties een stuk gezakt zullen zijn. De mate waarin dit gebeurt is, wat de student betreft, afhankelijk van bijvoorbeeld de mate waarin hij de stof louter gememoriseerd en niet begripsmatig verwerkt heeft; wat het tentamen betreft is het afhankelijk van de mate waarin eenvoudige feitenkennis gevraagd wordt dan wel hogere cognitieve vaardigheden als "inzicht". Leer psychologisch onderzoek toont aan (zie ook Cook, 1951) dat feitenkennis snel verdwenen blijkt te zijn bij onverwachte toetsing een tijd later, maar dat verworven inzichten in samenhangen e.d. vaak nog onverminderd aanwezig zijn. Het is jammer dat juist het voorgestelde propedeutisch examen hoogst waarschijnlijk primair aanspraak zal doen op geheugenwerk en niet op inzichtelijk met de stof kunnen omgaan. (Zie op dit punt ook Cronbach's artikel over validiteit in Thorndike

(1970)).

Veranderingen in resultaten zoals juist beschreven zijn uiterst ongewenst wanneer het er om gaat iets te weten te komen over de stabiele eigenschappen van de student die van belang zullen zijn voor de wijze waarop hij zich in de verdere studie zal gedragen. We zullen op zijn minst inhoudelijke examenprocedures moeten ontwerpen die in veel mindere mate op dit punt te kritiseren zijn dan de huidige, doen we dat niet dan laten we een belangrijke bron van foutieve selectiebeslissingen bestaan.

19

DE CRITERIUM MAATSTAVEN (Het "peil")

In het voorgaande hebben we het propedeutisch examen als selectie-instrument besproken, met al wat er aan consequenties aan verbonden is. We hebben tot nog toe a.h.w. alleen over de werkwijze gesproken, en nog niet over de rationale, we hebben immers het "geschiktheids"-probleem nog even uitgesteld. Wat we tot nu toe "geschiktheid" hebben genoemd dopen we, in overeenstemming met de literatuur over selectie, het **criterium**, of de maatstaf, eventueel nog dubbel op: criterium maatstaf.

Hiermee verwijderen we ons van het dagelijks spraakgebruik waarin "geschiktheid" gezien wordt als een "eigenschap" die een student

"bezit" of niet "bezit", aan het begin zowel als gedurende de studie. Het criterium daarentegen is een omschrijving van het (lieft meetbare) einddoel van het onderwijs; een einddoel dat de student, gesteund door het hem geboden onderwijs, in meerdere of mindere mate voor zichzelf zal realiseren (tenzij hij /zij om de een of andere reden vóór die tijd met de studie is opgehouden). In plaats van over een eigenschap van de student spreken we nu over een doelstellingenpakket van het onderwijs, dat voor een bepaalde student in een bepaalde mate bereikbaar blijkt.

In het klassieke test-model gaan we er van uit dat we de selectie zó inrichten dat we die studenten aannemen die uiteindelijk de beste prestaties zullen leveren, afgemeten aan de voor dat bepaalde stuk onderwijs geformuleerde doelstellingen (of een dergelijke werkwijze zonder meer gevolgd mag worden zal onderwerp van discussie zijn in een volgend hoofdstuk).

De vraag is dus hoe we over die eindprestaties gegevens kunnen verkrijgen, en welke waarde we aan die gegevens kunnen toekennen. Zoiets als 'het' criterium bestaat natuurlijk niet. Allereerst kunnen we onderscheid maken tussen criteria die te maken hebben met de studie bijvoorbeeld studieduur of gemiddeld

eindexamen-cijfer, en criteria die te maken hebben met de wijze waarop de vroegere student zich in zijn beroepssituatie gedraagt, bijvoorbeeld aantal publicaties, maatschappelijke positie, plezier dat hij/zij in het werk heeft e.d.. We kunnen de criteria die betrekking hebben op de periode ná de studie maar beter vergeten, niet omdat ze onbelangrijk zouden zijn, integendeel zelfs, maar omdat ze in hoge mate ongrijpbaar zijn, en alleen zeer kostbaar onderzoek daarover de nodige informatie zal kunnen verschaffen. Wanneer we ons beperken tot criteria die rechtstreeks met de studie te maken hebben, moeten we ons daarbij steeds blijven realiseren dat we ons daarmee in feite slechts met tussenliggende criteria bezighouden, de uiteindelijke criteria liggen tenslotte altijd in de beroepsuitoefening.

De studie-criteria moeten afgeleid worden uit de doelstellingen van het onderwijs, de meer inhoudelijke doelstellingen zowel als de meer fundamentele, principiële, maatschappelijke, of hoe men ze noemen wil. Wanneer het niet lukt om de doorgaans slechts zeer vaag omschreven doelstellingen, waarmee men nog alle kanten uitkan, scherper en uitgebreider te formuleren, zodat in de gekozen doelstellingen precies datgene geformuleerd is wat men

d.m.v. het onderwijs bij de studenten wil bereiken, zal het niet mogelijk zijn criterium prestaties van de studenten tijdens of als afsluiting van de studie vast te stellen. We hebben die doelstellingen absoluut nodig, omdat ons criterium precies het onderwijs doel moet representeren, niets meer, en ook zeker niets minder.

Afhankelijk van de doelstellingen zullen we verschillende criteria kunnen onderscheiden. Te noemen vallen bijvoorbeeld: studieduur, kwaliteit van zelfstandig onderzoek, examencijfers, gebleken maatschappelijk verantwoordelijkheidsbesef, houding in sociale omgeving, etc. Dit zijn stuk voor stuk zaken die moeilijk of niet kwantificeerbaar zijn, of, in het geval van de cijfers en de studieduur, waar we grote moeilijkheden bij het interpreteren van de cijfers zullen hebben. Maar het is in principe denkbaar dat we voor iedere student aan het eind van de studie een serie gegevens hebben met betrekking op de relevante criteria.

Zijn deze gegevens echter voor verschillende studenten met elkaar vergelijkbaar? In het hoger onderwijs zeker niet, want iedere student heeft a.h.w. zijn eigen persoonlijke studieprogramma doorlopen, heeft onderzoek gedaan over onderwerpen die hem interesseerden, heeft deelgenomen aan een

groot aantal, telkens verschillend samengestelde groepen, heeft speciaal voor hem samengestelde literatuurlijsten als tentamen opdrachten gehad etc.

- 21 -

Waar we bij het propedeutisch examen nog zo makkelijk de resultaten van verschillende studenten met elkaar konden vergelijken, zij het dat we de nodige kanttekeningen daarbij moesten maken, blijkt het relatief nagaan van studieprestaties op de daarvoor gehanteerde criteria niet meer mogelijk, tenzij we met beoordelaars gaan werken, met alle overweldigende bezwaren van dien. We blijven dan zitten met heel interessante gegevens voor iedere afzonderlijke student, maar we kunnen ze nauwelijks met elkaar vergelijken.

Als illustratie de conclusie van een onderzoek door Kelly en Fiske, zoals geciteerd en vertaald door Willems (1959): *"Een bevredigend enkelvoudig criterium voor succes bij de opleiding voor of bij het beoefenen van de klinische psychologie is er niet. De universitaire stafleden en de klinische psychologen, die de supervisie hadden, vertonen grote individuele verschillen in hun opvattingen over wat een succesvolle academische opleiding is en over wat vakbekwaamheid is."* Willems besluit een vrij uitvoerige bespreking van het criterium

probleem met o.a. de volgende opmerking: *"De universiteit moet de doeleinden van de opleiding duidelijk aangeven en expliciteren in hoeverre deze doeleinden terug te vinden zijn in de eisen welke bij het doctoraalexamen, of equivalent daarvan, gesteld worden"*.

Met het voorgaande betoog is tevens duidelijk geworden dat degenen die zo makkelijk praten over het 'peil' en het 'niveau' van de opleiding woorden in de mond nemen die zij heel moeilijk door concrete demonstratie waar zullen kunnen maken.

Rest ons ten aanzien van de criteria nog de belangrijke opgave na te gaan waardoor de uiteindelijke studiestatistiek zoals afgemeten aan deze criteria, bepaald kunnen zijn. Het ligt voor de hand hierbij te zoeken naar kwaliteiten van het onderwijs, naar kwaliteiten van de student, en naar hun onderlinge samenhang. Ook kan de gehanteerde meetprocedure van invloed zijn, omdat bij verschillende methoden verschillende resultaten zullen blijken. De geformuleerde doelstellingen vormen tevens een beperking, prestaties die daar buiten vallen ta zullen niet meetellen, omdat we ze niet meten. Onredelijk geformuleerde doelstellingen kunnen ten onrechte de indruk geven dat er iets mis is met het

onderwijs of de studenten, waar in feite de doelen fout, of te hoog of te laag gesteld zijn. Tenslotte zijn de studenten die de eindstreep bereiken het resultaat van een langdurig selectieproces, deze selectie heeft gevolgen voor de meting van de criteria, o.a. ten aanzien van de score-variantie, maar waarschijnlijk ook op meer systematische wijze afhankelijk van de variabelen die samenhangen met het vroegtijdig moeten beëindigen van een studie.

Allereerst de kwaliteit van het onderwijs. Kolthoff, onderwijs onderzoeker aan de Universiteit van Amsterdam, in de Haagse Post van 6 november jl.: *"Het grootste bezwaar van ons wetenschappelijk onderwijs is niet dat het zo duur, zo inefficiënt en zo lang is, maar dat het zo slecht is"*. Dit is een opmerking die relevant is voor het gehele selectie-probleem. Op dit moment interesseert ons alleen de onderwijskwaliteit. Deze kwaliteit moet, onder de gegeven omstandigheden, zo groot mogelijk zijn. Wanneer dit niet het geval is, en de opmerking van Kolthoff stelt onomwonden het onderwijs in gebreke, moeten we vele twijfels koesteren ten aanzien van de onder deze omstandigheden verkregen criterium gegevens.

Wanneer de nodige garanties t.a.v. de kwaliteit van het onderwijs niet gegeven kunnen

worden, zullen we niet weten of prestaties van studenten voornamelijk aan henzelf ofwel aan tekortkomingen in het onderwijs toegeschrevene moeten worden. Het is op zich al een moeilijk probleem om de individuele- en de onderwijsbijdrage uit elkaar [te halen], en dat wordt alleen maar moeilijker gemaakt naarmate het onderwijs zelf slechter is, meer op een random proces gaat lijken.

De kwaliteiten van de studenten die de eindstreep halen zullen uiteraard grote verschillen tonen. Het zal echter ontzettend moeilijk worden om aan deze verschillen oordelen als "beter" of "slechter" te verbinden, want daartoe moeten we zeer verschillende criteriumscores combineren tot één score op een dimensie goed-slecht (dit probleem komt in het volgende hoofdstuk ter sprake). En kwaliteiten die betrekking hebben op creativiteit, kritische houding, zelfstandigheid, verantwoordelijkheidsbesef etc. zijn erg moeilijk te concretiseren. Dat betekent dat we misschien wel kunnen theoretiseren over de kwaliteiten van de student die bepalend zijn voor studie-criterium-prestaties, maar dat het aanwijzen van empirische

- 23 -

samenhangen erg moeilijk zal worden. En dát is een frustrerende omstandigheid voor degenen

die een aankomende studenten-populatie moeten selekteren, o.a. juist rekening houdend met deze kwaliteiten.

Dat er belangrijke onderlinge effecten zijn tussen kwaliteiten van studenten en die van het onderwijs, behoeft nauwelijks toelichting. Ook hier is echter de grote moeilijkheid een inzicht te krijgen in de aard, de omvang en de consequenties van het onderlinge beïnvloedingsproces tussen de student en zijn onderwijs-omgeving.

Dat bij verschillende meet-procedures andere verdelingen van de resultaten zullen optreden, ligt voor de hand omdat er niet één criterium is voor zoiets als succesvol beëindigen van de studie, waar we vele criteria kunnen onderscheiden die vaak met elkaar weinig verband houden. Het kiezen van een meet-procedure hoort dan ook op basis van zorgvuldige overwegingen te gebeuren, met daarbij als bedenking dat een zekere mate van willekeurigheid onvermijdelijk zal zijn. Als basis voor die overwegingen zullen de geformuleerde doelstellingen kunnen dienen, de hele gang van zaken is zelfs te beschouwen als operationalisatie van de doelstellingen, het concretiseren in termen van gedragingen van studenten van wat daarvoor nog slechts

verbale abstracties waren. Aan een dergelijke ope-rationalisatie van onderwijsdoelstellingen kleven echter bezwaren waar men wél rekening mee zal moeten houden (Cronbach, 1971).

Tenslotte, waar het zo is dat we in het begin van de studie een prijs zetten op volgzaamheid, geheugenwerk, onzelfstandigheid, onderlinge wedijver etc., kunnen we niet verwachten dat degenen die de studie doorzetten aan het eind daarvan gekenmerkt kunnen worden door kritisch inzicht, creativiteit, zelfstandigheid, vermogen tot samenwerking etc. In de mate waarin dat ondanks alles tóch het geval is, moeten we verwachten op dit punt véél betere resultaten te kunnen behalen door in het begin van het onderwijs aan deze zaken meer aandacht te schenken en niet de verkeerde mensen uit te selecteren of het leven zuur te maken.

24

VOOSPELENDE GELDIGHEID VAN DE SELEKTIE

We hebben nu een fors stuk voorbereidend werk achter de rug, en zijn min of meer klaar voor het beantwoorden van de vraag of er verband bestaat of bestaan kan tussen scores op het propedeutisch examen en criterium scores aan het eind van de studie. Daarna zullen we

moeten onderzoeken wat het rendement van de selectie-procedure is, gegeven bekendheid van de voorspellende geldigheid.

Bij selectie à la Posthumus werken we in feite met een zeer complex model: enerzijds worden mensen via een ingewikkeld procedé van herkansingen afgewezen, anderzijds bieden we degenen die slagen een grote keuzevrijheid aan t.a.v. vormgeving aan de studie. Er wordt niet gewerkt met een bepaald aantal te vullen plaatsen, of met aannemingspercentages (Tenminste, het is nergens zo geformuleerd. In de praktijk kunnen natuurlijk wel andere dingen blijken te gebeuren). Omdat de geschetste situatie te complex is voor een analyse, gaan we er van uit dat er slechts twee mogelijkheden zijn bij een definitieve selectiebeslissing: aannemen of afwijzen. Op het feit dat mensen die aangenomen worden niet dezelfde "behandeling" krijgen, en dat er gewerkt wordt met herkansingen, komen we misschien later terug omdat er wel degelijk consequenties uit voortvloeien t.a.v. voorspellende geldigheid en rendement.

Wat is voorspellende geldigheid? Het gaat om de mate waarin scores op het propedeutisch examen iets zeggen over scores die later bereikt worden op de criterium toetsen aan het

eind van de studie. Hoe groter dit verband is, meestal uitgedrukt als correlatie-coëfficiënt, hoe groter de voorspellende geldigheid. Voorspellende geldigheid zullen we niet beschouwen als eigenschap van het examen, maar van de **beslissingsregel** die we hanteren t.a.v. zakken en slagen. Dit betekent dat we niet moeten kijken naar de examen-cijfers, maar slechts naar het gegeven voldoende-onvoldoende voor het gehele examen als dat inderdaad samenvalt met de beslissing afwijzen-aannemen. Dit is zeker geen onbelangrijk punt, het is namelijk mogelijk dat we een grote voorspellende geldigheid

25

vinden wanneer alle tentamencijfers bij de berekening gebruikt worden, en toch een lage voorspellende geldigheid vinden wanneer uitgegaan wordt van het gegeven voldoende-onvoldoende. Gegevens van deze aard zijn bijvoorbeeld te vinden in de dissertatie van Mellenbergh over studietoetsen (Psychol.Lab., Universiteit van Amsterdam, 1971) waar hij een zeer grote betrouwbaarheid van beoordeling van tentamen vragen vindt tussen een aantal beoordelaars, terwijl slechts over iets meer dan de helft van de studenten overeenstemming tussen de beoordelaars bestond of ze gezakt dan wel geslaagd waren! (Mellenbergh heeft dit

punt geheel over het hoofd gezien).

Welnu, voor het vinden van de voorspellende geldigheid van onze beslissingsregel (= selectie-procedure), moeten we logischerwijze een experiment opzetten waarbij na het examen alle studenten de kans krijgen om de studie af te maken, wat voor examenresultaten ze ook gehad mogen hebben.

"Bij onderzoek naar selectie vereist de experimentele opzet dat we iedereen aannemen, ongezien de examencijfers. (...) De verdere studie-resultaten worden verzameld. Deze waargenomen studie-resultaten zijn een afspiegeling van de doelstellingen van de onder-wijsinstelling of van de student".
(Cronbach, 1971, blz.484).

Dit betekent in de praktijk dat we, voor iedere afzonderlijke studierichting een totale groep aankomende eerstejaars zonder enige vorm van selectie tot aan het doctoraal-examen moeten proberen te brengen. Studenten die op bepaalde momenten moeilijkheden bij de studie ondervinden, die in principe door goede begeleiding of beter onderwijs te voorkomen zijn, moeten zoveel mogelijk overreed worden door te gaan met de studie. Op zich is een dergelijke gang van zaken al opzienbarend, en het valt te voorzien dat ook de resultaten opzienbarend zullen zijn; nl. dat vele van

"ongeschikt" bevonden studenten die we toch door laten studeren zullen blijken goede criterium-prestaties te behalen, en omgekeerd. En dit is precies de soort informatie die we nodig hebben; zonder een dergelijk, ongetwijfeld kostbaar,

- 26 -

project zullen we geen voorspellende geldigheid van de selectie-procedure kunnen bepalen. *)

Zonder inzicht in de voorspellende geldigheid weten we niet wat we doen, en in dergelijke gevallen kun je maar beter direct overgaan op lotings-procedures. Als er selectieprocedures voorgesteld worden, moet aangetoond worden dat ze "beter" zijn dan lotings-procedures; dat is het hele probleem waar we ons in dit rapport mee bezig houden.

Wanneer we tenslotte examen- zowel als criterium-gegevens hebben verzameld (en in ons geval zal daar tenminste **een jaar of zes**, mee gemoeid zijn) moet onderzocht worden welke beslissingsregel, toepasbaar op de examenresultaten, een zo gunstig mogelijk resultaat zoals dat aan de criteria is afgemeten, oplevert. Wanneer we van meervoudige scores gebruik maken, moeten we zoeken naar die "gewogen" combinatie van afzonderlijke scores die een "zo goed mogelijke" beslissingsregel

oplevert. In dit laatste geval moeten we echter het hele experiment nóg eens uitvoeren op een andere groep om te controleren of de specifieke wegings-combinatie die we gekozen hebben ook geldig is voor een andere groep. Vaak namelijk vinden we bij gebruikmaking van multi-variatie scores en methoden dat oorspronkelijke uitkomsten specifiek zijn voor de gebruikte groep proefpersonen, en voor andere groepen niet of in mindere mate gelden. In plaats van een beslissingsregel op empirische gegevens te baseren, kunnen we deze ook à priori formuleren. Dit ontslaat ons echter geenszins van de verplichting om het experimenteel onderzoek uit te voeren, omdat ook van deze á priori regel de geldigheid aangetoond moet worden (zie ook Cronbach, 1971, blz.485).

*) zgn. "concurrente validering" geeft geen enkele zekerheid over de feitelijke voorspellende geldigheid, en deze procedure die veel goedkoper zou zijn, kan niet gehanteerd worden. "Such concurrent validation must not be confused with a true follow-up study" (Cronbach, 1971, blz.484).

bepaalde studierichting) vereist een behoorlijk grote experimenterale groep en een verdedigbaar criterium - vereisten waaraan wel eens moeilijk tegemoet kan worden gekomen" (Cronbach, 1971, blz.486).

Bij het uitvoeren van de geschetste experimentele procedure moeten we tevens allerlei effecten proberen uit te schakelen of te controleren die verband houden met verwachtingen van docenten over de prestaties van hun studenten. Talloze onderzoeken hebben de invloed van dergelijke verwachtingen van de docent op het objectieve prestatieniveau van hun leerlingen aangetoond (zie o.a. Willems, 1959 voor een bespreking).

Dit effect speelt natuurlijk ook in de gewone onderwijssituatie, waar het al betreurenswaardig genoeg is. In onze experimentele opzet moeten we al hele speciale waarborgen creëren dat propedeuse resultaten die door studenten behaald zijn onbekend blijven aan de docenten, zodat propedeuse en criterium resultaten zoveel mogelijk experimenteel onafhankelijk van elkaar blijven.

"Empirische bepaling van de geldigheid is onvermijdelijk retro-spectief, terwijl het gebruik van de test toch altijd prospectief is." (Cronbach, 1971).

Vooraf in onze situatie betekent dit dat we na al onze moeite uiteindelijk een gegeven over de voorspellende geldigheid hebben verkregen waarvan we niet weten in hoeverre we daarvan ook uit mogen gaan bij nieuwe te selekteren groepen. Het duurt tenminste een jaar of zes voordat het onderzoek gegevens oplevert op basis waarvan een voorspellende geldigheid te berekenen is. Dat betekent dat op het moment dat we hiervan praktisch gebruik willen gaan maken, er vele veranderingen zijn opgetreden in de situatie zoals die zes jaar geleden was: het onderwijsprogramma is vaak nogal ingrijpend veranderd, de inhoud van het vak is veranderd, er zijn mutaties in het docentenkorps opgetreden, en de groep studenten die aankomt moet op vaak andere wijze gekarakteriseerd worden als de oorspronkelijke proefgroep."

"Een onderzoek waarin (studie)succes voorspeld wordt met behulp van een statistische formule krijgt haar duidelijke betekenis wanneer de formule ontwikkeld wordt op de plaats waar zij ook toegepast zal worden (de betreffende studierichting) en deze situatie voldoende stabiel is dat de resultaten representatief zijn voor

28

hetgeen in volgende jaren zal gebeuren. Slechts

wanneer het aanbod van studenten en het onderwijs-programma goeddeels hetzelfde karakter behouden, is het aannemelijk dat de resultaten direct toepasbaar blijven" (Cronbach, 1971, blz.485).

De grote flexibiliteit waardoor wetenschappelijk onderwijs altijd gekenmerkt zal moeten zijn, legt nog een extra zwaar accent op de gesignaleerde tegenstrijdigheid tussen de genomen selectie-beslissingen en de ondertussen verouderde gegevens op basis waarvan we de beslissingeregels hebben geformuleerd.

Het is nu tijd ons te bezinnen op de mogelijke consequenties van het niet rekening houden met aan de beroepspraktijk ontleende criteria. Deze uiteindelijke criteria zijn moeilijker te voorspellen dan studie-criteria omdat ze veel verder in de tijd verwijderd zijn van de selectie-situatie, omdat behoorlijke metingen moeilijk verkregen kunnen worden etc.. Dit zijn evenzovele redenen waarom de voorspellende geldigheid die we vinden t.a.v. studie-criteria een overschatting zal zijn van de voorspellende geldigheid t.a.v. succes in de beroepssituatie. En deze overschatting wordt nog sterker naarmate die beroepssituatie sterker verschilt van het geprononceerd verbale karakter

waardoor onderwijs gekenmerkt wordt: *"Eén van de redenen waarom studie-criteria beter voorspelbaar zijn (dan beroepsriteria) is dat zij veelal afhanke-lijk zijn van verbaal gedrag (taalbeheersing) in school en op examens, wat makkelijk voorspelbaar is, terwijl succes in het beroep afhankelijk is van niet-verbale kwaliteiten die moeilijk te meten zijn. Maar al te vaak worden mensen uit sociaal minder bedeelde bevolkingsgroepen in het nadeel gesteld door de gehanteerde studiecriteriën; en een test die een criterium "geldig" voorspelt, zou in dit opzicht een onjuiste grond voor (selektieve) beslissingen zijn."* (Cronbach, 1971, blz.487)

Het is een algemene ervaring bij selectie-onderzoek dat discrepanties als de juist genoemde tussen uiteindelijke (beroeps-) criteria en tussenliggende (studie)-criteria optreden. Een waarschuwing van Cronbach: *"Bij het uitvoeren van selectie-onderzoek moeten we voortdurend de verleiding weerstaan om ons te concentreren op criteria die makkelijk te voorspellen zijn. We moeten onze aandacht richten op die criteria die we het belangrijkste vinden"*. (1971, blz.490). Het gevaar dat hier gesignaleerd wordt

29

is bijzonder groot. En wanneer we er niet speciaal

op bedacht zijn, kan het nog sterk vergroot worden door het verschijnsel dat ook in de Nederlandse literatuur al bekendheid geniet als "restriction of range". Voor ons is het volgende aspect hiervan van belang: de aankomende student heeft een continue selectieproces van 12 jaar of meer achter de rug. Niettegenstaande de grote mate van willekeurigheid van deze selectie, die immers meer een sociale selectie vormt, heeft zij er toch in geresulteerd dat op een aantal eigenschappen de groep aankomende studenten een extreme groep is, vergeleken met de totale leeftijdsgroep. Dit is o.a. het geval t.a.v. intelligentie, en vooral al dié variabelen die samenhangen met de "overlevingskans" in het onderwijs. Dus ook in het hoger onderwijs. Met andere woorden: juist op een aantal variabelen die bij uitstek geschikt zouden zijn om te gebruiken bij de selectie-procedure, juist op dié variabelen **zijn de studenten al geselecteerd**. En zowel theoretisch als empirisch is aantoonbaar dat in een dergelijke situatie de betreffende variabelen moeilijk of niet bruikbaar zijn bij verdere selectie. Dit verklaart een belangrijk deel van de bijna onoverkomenlijke moeilijkheden bij het ontwerpen van een rationele selectie-procedure voor het hoger onderwijs. Iedereen

is bij wijze van spreken al geschikt omdat hij /zij een eindexamen-papiertje bezit. Wanneer we, ondanks dit gegeven, tóch een selectie-procedure ontwikkelen, lopen we daarbij groot gevaar in de door Cronbach genoemde val te lopen. Immers, op veelal werkelijk relevante eigenschappen verschillen de aankomende studenten niet genoeg meer om er op te kunnen selekteren, in welk geval het gevaar erg groot is dat er dan geselecteerd gaat worden op variabelen waarop er wél grote verschillen bestaan tussen de studenten, **maar die voor het onderwijs minder belangrijk zijn**. In feite zouden we dan met een selectie bezig zijn die neerkomt op willekeur.

Dan rest ons nu nog het probleem, waaraan we in het voorgaande stilzwijgend voorbijgingen, dat er op de een of andere manier verband gelegd moet worden tussen een aantal examen-scores (bijvoorbeeld de verschillende tentamens waaruit dat examen bestaat) en een verzameling scores op al dié verschillende criteria die we geformuleerd hebben.

30

Het is waarschijnlijk dat deze verschillende criteria tamelijk onafhankelijk t.a.v. elkaar zijn (precies de reden waarom we ze niet makkelijk kunnen combineren tot één enkele criterium score). Ook met de examen-scores zal dit (in mindere

mate) het geval zijn. Maar niet alleen kunnen scores op de criteria weinig verband (lage intercorrelaties) met elkaar hebben, ook blijken zij vaak met verschillende variabelen (onderdelen) van het voorspellende examen af te hangen. *"De tests die het ene resultaat voorspellen zijn vaak niet degene die een ander resultaat voorspellen, en een beslissingsregel waarin een bepaald resultaat gemaximaliseerd wordt zal mensen kunnen afwijzen die op een ander criterium uitmunten"*. (Cronbach, 1971). Natuurlijk kunnen we alle verbanden tussen de verschillende scores analyseren, maar dat levert nog geen beslissingsregel op. Er zijn ook geen statistische methoden die ons hierbij kunnen helpen (ook canonische correlatie niet, zie Cronbach, 1971, blz.489-490) omdat *"de keuze van de combinatieregels afhangt van de doelen of de waarden van degene die beslist, d.w.z. de doelen of waarden van het instituut dat hij vertegenwoordigt"*. (Cronbach, 1971).

Het zal veel moeite en veel vergadertijd kosten om tot dergelijke combinatie-regels te komen. Het is echter noodzakelijk er aandacht aan te besteden omdat op andere wijze verkregen beslissingsregels in elk geval een impliciete weging van examen- en criteriumscores behelzen. Dan zouden bijv. de minder

belangrijke, maar makkelijk voorspelbare, criteria een te groot gewicht kunnen krijgen.

- 31 -

HET RENDEMENT VAN DE SELEKTIE PROCEDURE.

We weten nog steeds niet of we met onze selectie-procedure nu iets gewonnen hebben, of dat net zo goed geloot had kunnen worden. Om deze vraag te kunnen beantwoorden, is het gegeven van de voorspellende geldigheid nodig, zoals dat verkregen kan worden door het uitvoeren van een onderzoek zoals dat in het voorgaande beschreven is. De voorspellende geldigheid wordt meestal uitgedrukt in een correlatie-coëfficiënt, de produkt-moment correlatie tussen selectie en criterium score. Deze is bij perfecte voorspellende waarde gelijk aan 1.00, bij totaal ontbreken van voorspellende waarde is ze gelijk aan 0.

Welke waarde van deze zgn. geldigheids coëfficiënt is te verwachten? Cronbach (1970): *"Het is ongebruikelijk om geldigheids-coëfficiënten te vinden hoger dan .60, hoewel dat nog lang geen perfecte voorspelling is. Hoewel we liever hogere coëfficiënten zouden hebben, is toch iedere positieve correlatie een aanwijzing dat voorspellingen op basis van de toets nauwkeuriger zijn dan bij raden. Of een*

geldigheidscoëfficiënt hoog genoeg is om voorspelling op basis van de toets te rechtvaardigen, hangt af van praktische overwegingen zoals de noodzaak van verbeterde voorspellingen, de kosten verbonden aan de toetsings-procedure, en de kosten en de geldigheid van de al in gebruik zijnde selectie-methoden. Op de vraag "Wat is een goede geldigheids-coëfficiënt", is het enige zinvolle antwoord: "De beste die je kunt krijgen". Als een criterium voorspeld kan worden met slechts een geldigheid van .20, dan zou de toets nog steeds een op prijs te stellen praktische bijdrage kunnen leveren. Natuurlijk moet deze bijdrage groter zijn wanneer het gaat om een kostbare selectieprocedure dan bij een relatief goedkope". (blz,135).*

In de situatie die zich voordoet bij selectie voor hoger onderwijs in Nederland mogen geen geldigheids-coëfficiënten hoger dan .30 of .40 verwacht worden, als het ons al mogelijk blijkt om ook maar enige geldigheid aan te tonen. (Willems 1959 en 1964, Hazewinkel 1971, Van Naerssen 1970). Wanneer op basis van onderzoek een coëfficiënt gevonden wordt om .30 bijvoorbeeld, moet daarbij bedacht worden dat we werken met criteria die ook betrekking hebben op studiesituaties, en indachtig

Cronbach's waarschuwing dat onderwijs-criteria makkelijker toorspelbaar zijn dan

32

praktijk criteria, moet deze waarde van .30 heel voorzichtig geïnterpreteerd worden.

Welke praktische betekenis heeft nu een bepaalde, gegeven, geldigheids-coëfficiënt? Allereerst kan op basis daarvan berekend worden hoeveel foutieve toelatings- en afwijzingsbeslissingen genomen worden om deze berekeningen, voor verschillende afwijzings-percentages, te kunnen uitvoeren moet een aanname gemaakt worden over het aantal "geschikte" studenten in de aankomende groep. Daarbij kan gebruik gemaakt worden van door empirisch onderzoek verkregen gegevens om op basis daarvan voor een bepaalde nieuwe groep een schatting te maken. Ook kan een à priori schatting gemaakt worden die redelijk lijkt. Laten we dan veronderstellen dat 80% van de aankomende eerstejaars de studie met succes en binnen redelijke tijd kan afsluiten. Dit is een aanname waar ook Hazewinkel (1971) van uitgaat bij door hem uitgevoerde berekeningen waarvan we hier gebruik zullen maken.

RENDEMENT TABEL. Uitgaande van 80% geschikten onder de aankomende studenten, en voorspellende geldigheid van .40.

afwijzings %	% foutief toegelatenen	% foutief afgewezenen	totaal fouten %
0	20	..	20
10	15,3	5,3	20,7
20	12	12	24

Het aantal ongeschikten in de aangenomen groep neemt af naarmate een hoger afwijzingspercentage gehanteerd wordt, het aantal mensen dat afgewezen wordt maar die potentieel de studie goed hadden kunnen volbrengen, neemt echter sterk toe. Wat er gebeurt wanneer er geheel willekeurig geselecteerd wordt, bijvoorbeeld d.m.v. een lotings-procedure, wordt geïllustreerd door de resultaten vermeld in onderstaande tabel.

33

RENDEMENT TABEL. Uitgaande van 80% geschikten onder de aankomende studenten, en selectie d.m.v. loting. (tussen haakjes de overeenkomstige cijfers uit de eerste tabel).

afwijzings %	% foutief toegelatenen	% foutief afgewezenen	totaal fouten %
0	20	..	20
10	18 (15,3)	8 (5,3)	26 (20,7)
20	16 (12)	16 (1)	32 (24)

Bij gebruik van een selectieprocedure met voorspellende geldigheid van .30, een à priori

geschiedenis van 80%, en een afwijzingspercentage van 20%, worden 8% minder foutieve beslissingen gemaakt dan bij een lotingsprocedure!

We moeten hierbij wel bedenken dat het argument dat loting resulteert in een slechter resultaat omdat enkele zeer goede studenten afgewezen worden en enkele zeer slechte aangenomen, terwijl bij selectie de verschillen in capaciteiten tussen onterecht aangenomenen en afgewezenen kleiner zijn, dat dit argument slechts voor een klein deel opgaat. De mate waarin dit argument geldig is wordt namelijk ook bepaald door de voorspellende geldigheid van de selectieprocedure: hoe kleiner deze is, hoe minder deze redenering opgaat. Een validiteit van .40 maakt deze redenering erg zwak, omdat de selectie-resultaten dan slechts 16% van de variantie in criterium scores, d.w.z. het onderscheid tussen wie goede en wie slechte studenten gebleken zijn, verklaart.

Op basis van bovenstaande gegevens moet geconcludeerd worden dat, wanneer we alléén letten op het gegeven van de voorspellende geldigheid en geen rekening houden met bijvoorbeeld de kosten van de selectieprocedure, er beter geselecteerd dan geloot kan worden, hoewel de winst die behaald kan worden teleurstellend klein is. Bovendien moet

bedacht worden dat stilzwijgend aangenomen is dat een beslissing die iemand onterecht afwijst even zwaar weegt als een beslissing die iemand onterecht aanneemt. Wanneer we willen proberen zo weinig mogelijk foutieve afwijzings-beslissingen te maken, omdat we liever mensen onterecht aannemen dan anderen onterecht af te wijzen, zal ook daarmee in de te volgen selectie-procedure rekening gehouden moeten worden.

34

Uit het bovenstaande is tevens gebleken dat voor het rendement van de selectieprocedure ook van belang is het selectie-percentages, het percentage mensen die we aannemen of afwijzen. Wanneer we dit percentage vrijelijk kunnen bepalen, dan kan op dit punt een "optimale" strategie gevolgd worden. (zie Van Naerssen, 1970, waaraan we het volgende numerieke voorbeeld ontleen).

Wanneer bijvoorbeeld 10% van de aankomende studenten ongeschikt is voor de studie (een verdedigbare aanname), en we beschikken over een selectief examen met een voorspellende geldigheid van .50 (en dat is behoorlijk hoog!) dan is het onder deze condities de beste strategie om 5 van de 1000 studenten af te wijzen! In dit geval zouden de kosten van de selectieprocedure zeker niet

opwegen tegen de winst die er mee behaald werd.

(Hierbij moet opgemerkt worden dat Van Naerssen niet uitgaat van het aantal foutieve beslissingen, maar dat voor hem de selectie verhouding optimaal is, wanneer "de som van de utiliteiten van de geaccepteerde personen" maximaal is.)

Wanneer uit een groot aantal gegadigden slechts ongeveer de helft aangenomen zal worden, dan komen de genoemde cijfers veel gunstiger te liggen. Dan kan zelfs een examen met een geringe voorspellende waarde nog een winst opleveren, in die zin dat minder foutieve beslissingen worden genomen dan zonder het selectieve examen het geval zou zijn geweest. Voor het nederlandse hoger onderwijs is dit een situatie die zich echter bij hoge uitzondering zal voordoen.

Van Naerssen merkt in genoemd artikel over het nut van selectie op, m.b.t. het universitair onderwijs:

"Selectie kan alleen zin hebben als het percentage 'geschikten' bij het aanbod in de buurt van 50 ligt. Als bijvoorbeeld 80% van het aanbod op tijd de eindstreep haalt, dan heeft het (bij reële validiteiten) geen zin te selekteren op verwacht studiesucces. Als 80% van de afgestudeerden "zijn salaris waard is" dan heeft het geen zin

door studentenselektie naar kwaliteitsverbetering te streven."

35'

Zoals al herhaaldelijk in dit rapport opgemerkt, moet rekening gehouden worden met de kosten die verbonden zijn aan de selectieprocedure. Het is immers mogelijk dat deze zo hoog zijn dat ze niet opwegen tegen de vermindering van het aantal foutieve beslissingen die er mee te behalen zijn. In het voorbeeld ontleend aan van Naerssen is wel heel duidelijk dat de kosten van de selectie het verkregen voordeel al gauw overtreffen. Een groot deel van de kosten die verbonden zijn aan deze wijze van selekteren moet gezocht worden in schade die toegebracht wordt aan de kwaliteit van het onderwijs, waarover in een volgend hoofdstuk gesproken zal worden. Daarnaast is er nog de belangrijke overweging van psychische schade die door stress-situaties, verloren jaren en op onjuiste gronden verloren kansen toegebracht wordt aan individuele studenten; ook een maatschappelijk ongunstig effect is inherent aan het hanteren van competitieve methoden in opvoedings-situaties.

Meer in het oog lopende kosten zijn o.a. de volgende, waarvoor hier geen numerieke schattingen gegeven zullen worden:

- tijd door docenten en toets-specialisten besteed aan het maken, afnemen, en interpreteren van toetsen.
- financiële kosten die aan bovengenoemde activiteiten verbonden zijn
- de tijd die door studenten besteed wordt aan voorbereiding op en deelnemen aan tentamens en examens, die voor eigenlijke onderwijs-taken min of meer verloren is
- de zeer hoge kosten verbonden aan het uitvoeren van onderzoek naar de werkzaamheid van de selectie-procedure, zoals in voorgaande hoofdstukken beschreven. Een dergelijk onderzoek zou bovendien voor vrijwel iedere studierichting en universiteit afzonderlijk uitgevoerd moeten worden.
- de hoge administratieve kosten verbonden aan dergelijke toets-programma's
- bij selectie tijdens het onderwijs in plaats van vooraf komt daar nog bij de gedeeltelijk verloren gegane studietijd van de na één of meer jaren afgewezenen.

- 36

Bedenk dat deze kosten afgewogen moeten worden tegen bijvoorbeeld het behaalde voordeel in percentage foutieve beslissingen, of tegen behaalde extra utiliteit, zoals dat begrip door Cronbach en Gleser (1965) en Van Naerssen gehanteerd wordt. Afgezien nog vgl

de nadelige gevolgen die selectie heeft voor de kwaliteit van het onderwijs, lijken de genoemde financiële kosten een selectie-procedure tot een minder zinvolle aangelegenheid te maken dan veelal verondersteld wordt. Daarnaast moet niet vergeten worden dat selectie niet de enige mogelijke oplossing van het probleem is, immers, we kunnen ook proberen de onderwijssituatie op zinvolle wijze aan te passen aan een vergroot aanbod van studenten, eventueel binnen bepaalde financiële restricties. Zolang de mogelijkheden in dit opzicht nog niet ten volle uitgebuit zijn, moeten we erg voorzichtig zijn met het naïevelijk hanteren van selectie-procedures.

Een andere frustrerende factor bij het rendementsvraagstuk is de o.a. door Posthumus gesignaleerde wetmatigheid die optreedt in de percentages onvoldoendes-voldoendes die door docenten uitgedeeld worden.

Hofstee (1970) heeft in zijn inaugurele rede een treffende experimentele illustratie hiervan besproken. Het betrof een onderzoek van Aiken, waar in een aantal opeenvolgende jaren een steeds scherpere selectie voor een bepaalde studierichting gepleegd werd. Ondanks de hogere kwaliteit van de scherper geselecteerde groep (de selectie correleerde .70 met studiesucces!) bleven de studie-

resultaten ongeveer gelijk. De docenten bleken eenvoudig hun normen verschoven te hebben: *"de gemiddelde opleidingsresultaten van de strengst geselecteerde groep waren praktisch gelijk aan die van de ongeselecteerde studenten in het eerste onderzoeksjaar"*. Hofstee waarschuwt tegen de redenatie dat dan weliswaar het ge-middelde cijferresultaat hetzelfde is, maar dat dan toch ongetwijfeld het 'niveau' hoger zal liggen: Dit nml. is nog maar helemaal de vraag waarvoor Hofstee een aantal argumenten aanvoert.

37

Dergelijke gegevens doen vermoedens rijzen over de meer algemene geldigheid van het verschijnsel van "aanpassing van de beoordeling aan het niveau van de studenten". Hofstee: *"De schaarse gegevens wijzen erop, dat ook waar valide en aanvaardbare voorspellers van studie- of beroepsprestatie te vinden zijn, aanpassings- of feedback-verschijnselen kunnen optreden die het nuttig effect van selectie ongedaan maken. De schaal waarop die processen optreden is onbekend. Het zou bijzonder prematuur zijn, op grond van de hier uitgesproken theoretische verwachtingen te concluderen dat selectie altijd en overal zinloos is; wel mag de eis worden gesteld, dat het selectie-onderzoek zich gaat*

bezighouden met de vraag van het reële rendement".

Onze conclusie tot dusverre is dat selectie-procedures voor de hoger onderwijs situatie niet of nauwelijks zinvol zijn, d.w.z. dat de kosten niet opwegen tegen het bereikbare voordeel. Hierbij is nog geen rekening gehouden met de kritiek die op de geschetste gang van zaken bij het ontwikkelen van een behoorlijke selectie-procedure geleverd kan worden, en die de nodige extra vraagtekens zet bij de rechtvaardigheid van selectie. Ook de nadelige terugslag van het hanteren van selectieve methoden op de onderwijssituatie is nog niet in rekening gebracht. Beide punten vormen de volgende onderwerpen van beschouwing.

38

KRITIEK OP HET GESCHETSTE SELEKTIE MODEL.

Er kan nogal wat kritiek geleverd worden op het hanteren van een selectie-methode als juist beschreven. Een dergelijke kritiek is echter alleen van belang voor het aan de oppervlakte brengen van verborgen veronderstellingen die bij selectie in onderwijssituaties een rol kunnen spelen. Praktische gevolgen in de zin van

betere methoden vloeien er vrijwel niet uit voort omdat een theoretische fundering in de testtheorie voorlopig nog ontbreekt door de complexiteit van de aan de orde gestelde problemen. (Zie voor een en ander: Cronbach en Gleser, 1965). Daarom zal hier ook met een paar korte opmerkingen volstaan worden.

Een eerste punt betreft de gevolgen die, zowel voor de onderwijsinstelling als voor de betreffende personen, voortvloeien uit de selectie-beslissingen. Geprobeerd moet worden deze gevolgen zoveel mogelijk te inventariseren, omdat bij beslissingen juist deze mogelijke gevolgen van groot belang zijn. Niet alleen kan dan duidelijker worden of gehanteerde selectie-procedures wel in overeenstemming zijn met de verwachtingen die we er altijd van gehad hebben, bovendien wordt het dan meer mogelijk om een zeer bewuste strategie te kiezen. De moeilijkheid hierbij zal zijn om tot een onderlinge vergelijking te komen van zaken die vaak verschillende "waarden" vertegenwoordigen. Genoemd kunnen worden: voldoen aan maatschappelijke behoeften aan academici, wenselijke studieduur, wenselijke kwaliteiten bij de afgestudeerden, minimaliseren van de opleidingskosten, individualiseren van het

onderwijs, tegemoet komen aan ontplooiingsmogelijkheden die specifiek zijn voor individuele studenten etc. Dit vereist het afwegen tegen elkaar van zaken die niet objectief op hun waarde beoordeeld kunnen worden. Technieken hiervoor zijn misschien wel te ontwerpen als hulpmiddel voor een besluitvormingsproces (Bijv. Raiffa, Decision Analysis, 1968).

Een ander punt vloeit voort uit het groepsgecentreerde karakter van huidige selectietechnieken.

Wanneer we gaan selekteren op basis van de in vorige hoofdstukken geschetste procedure, neemt de onderwijs-instelling een groot aantal beslissingen: over iedere aankomende student wordt

- 39 -

beslist of hij /zij aangenomen of afgewezen wordt. Bij ieder van deze beslissingen worden dezelfde overwegingen of argumenten gehanteerd, een ieder wordt met dezelfde set van criteria vergeleken. Dit soort beslissingen wordt wel institutionele beslissingen genoemd, waarmee bedoeld is ze te contrasteren met individuele beslissingen. Cronbach & Gleser (1964, blz.130):

"Bij de individuele beslissingen is het niet mogelijk te kiezen tussen verschillende handelwijzen

(doorgaan of niet met de studie bijv.) behalve op grond van de waarden van degene op wie de beslissingen betrekking heeft. In een groep studenten die advies vragen over de te kiezen studierichting moet voor ieder student afzonderlijk de beslissing geëvalueerd worden op een verschillende waarde-schaal. Omdat de student een bepaalde keus maar éénmaal doet, is het evident onmogelijk om een strategie op te stellen die, door de bank genomen, de beste is, omdat een gemiddelde in dit geval niet zinvol te definiëren is. Een bepaald besluit moet geëvalueerd worden op basis van het verwacht resultaat en zijn waarde voor de betrokken persoon".

Het gaat hier niet alleen om het meer tot zijn recht laten komen van de autonomie, de keuzevrijheid van de individuele student, maar evenzeer om een maatschappelijk belang. Wanneer bij het nemen van selectiebeslissingen de persoonlijke waarden van de betrokken studenten niet betrokken worden, wordt de selectie wél een stuk "makkelijker" gemaakt voor degenen die ze uitvoeren, maar zullen ook meer foutieve beslissingen genomen worden omdat een belangrijke bron van extra informatie niet aangeboord is. Daarnaast zal de maatschappij meer mogelijkheden moeten scheppen voor de realiseerbaarheid van

individuele verlangens ten aanzien van opleidingsmogelijkheden. Dit laatste is vrij makkelijk te bereiken in het hoger onderwijs zonder dat daar direkt hoge kosten uit voort zouden vloeien, het lijkt echter gewenst om met een dergelijk model in het lager of voortgezet onderwijs te starten.

Gedachten als deze zijn o.a. te vinden bij Meuwese (1970), waar hij een door hem voorgesteld onderwijsmodel bespreekt: *"Het moet een systeem impliceren waarin elke leerling optimale kansen krijgt zijn individuele capaciteiten en talenten te ontwikkelen. Dat wil zeggen dat het systeem maximaal "adaptief" moet zijn dan de specifieke eigenaardigheden van elke leerling. In de klassieke terminologie: het moet geïndividualiseerd onderwijs mogelijk maken. In dit uitgangspunt zit het waardeoordeel opgesloten, dat elk individu recht heeft op maximale ontwikkeling van zijn vermogens, en dat die ontwikkeling niet beperkt mag worden door starheid van het onderwijssysteem. Ofwel: het onderwijs moet zich bij hem aanpassen, en niet hij zich aan het onderwijs".* Het laatste vinden we ook beschreven in Schwarz (in: Thorndike, 1971): *"Tests werden in feite gebruikt om de studenten aan te passen bij het gegeven onderwijsprogramma; en slechte*

studieresultaten werden gezien als het gevolg van verkeerde selectie-beslissingen. Maar in toenemende mate wordt het onderwijs nu aangepast aan de studenten op een individuele basis, met behulp van een scala van nieuwe onderwijstechnieken, En de slechte resultaten worden nu gezien als een tekortkoming van de gehanteerde onderwijs-methodieken, een inzicht dat weer leidt tot veranderingen in het onderwijs".

Dit standpunt is de laatste jaren ook in Nederland meer en meer te horen, en heeft grote implicaties voor de wijze waarop in het onderwijs geselecteerd kan worden.

41

KRITIEK OP SELEKTIE TIJDENS HET ONDERWIJS

In het voorgaande is al veelvuldig verband gelegd tussen selectie en onderwijs, de velerlei wijzen waarop beide elkaar beïnvloeden. Dit hoofdstuk zal alleen enkele nog niet besproken punten bevatten, en moet dus niet gezien worden als een volledige behandeling van de invloed die op de kwaliteit van het onderwijs uitgaat van de wijze waarop tijdens dat onderwijs geselecteerd wordt.

Posthumus signaleert een spanning tussen wat hij noemt de selectie en de onderwijsfunctie van de docent (De universiteit etc.blz.18). "*De wens*

om de onderwijstaak zo optimaal mogelijk te vervullen dat een begingeneratie in een vastgestelde tijd in haar geheel naar een einddoel kan worden gebracht komt in strijd met de functie van 'selectie' zoals die in de, in toenemende mate op diplomabezit berustende, nederlandse statushiërarchie aan het onderwijs is toegefallen. Iedere docent kent het probleem van de verdeling van zijn aandacht tussen de begaafde en de achterblijvende studenten. Iedere examinerator kent de dualistische strijd tussen welwillendheid tegenover de kandidaat en de vrees voor verlaging van het peil". Waarmee Posthumus wél een conflict signaleert, maar zich er vervolgens bij neerlegt dat het niet ten gunste van het onderwijs oplosbaar is. Het handhaven van het selectieve beoordelings-systeem zoals we dat nu kennen, en zoals dat met Posthumus gecontinueerd zal worden, heeft een ongewenste invloed op de kwaliteit van het onderwijs (Meuwese, 1970): "Het beoordelingssysteem is een van de krachtigste regelsystemen van, het onderwijs. In kwalificerende cijfers uitgedrukte oordelen bepalen het studeergedrag van de leerlingen, bepalen het rendement van het onderwijs, maken via het al of niet behalen van diploma's de maatschappelijke carrière van leerlingen uit. Het cijfersysteem kan er toe leiden dat

studenten en leerlingen niet meer gemotiveerd zijn een vak te beheersen, te leren denken of te leren kennis toe te passen, maar alles op alles te zetten om het cijfer te halen dat hen een stapje verder brengt op weg naar "het papiertje"."

42

Met deze effecten, die niet op zouden treden wanneer voorafgaand aan het onderwijs, of geheel niet, geselecteerd wordt, moet rekening gehouden worden bij de afweging van de voor- en nadelen van diverse selectie-procedures, loten of geheel niet selecteren. Voor zover nog niet eerder genoemd zullen een aantal belangrijke effecten besproken worden.

Voor de studie- en levenssituatie van de student maakt het enorm veel uit of hij moet werken onder de voortdurende dreiging er altijd nog uitgeschopt te kunnen worden, of dat hij zich in zijn studie gesteund weet door een staf van docenten die er allereerst op uit is hem in zijn studie vooruit te helpen. Het lijkt op het intrappen van een open deur dit telkens weer op te merken. Maar diezelfde deur zit bij Posthumus bijvoorbeeld, nog potdicht, hij wil immers een selectief propedeusejaar instellen, gevolgd door een doctoraal fase waarin de selectie "op de achtergrond" zou moeten treden. Nog afgezien van het ontbreken van

garanties in het wetsontwerp, dat selectie in de doctoraal-fase geen rol van betekenis meer zal kunnen spelen, blijkt uit de voorgestelde constructie toch een te grote aarzeling om uit verworven onderwilskundige inzichten de volledige conclusies te trekken. Posthumus komt halverwege, en verschanst zich vervolgens achter een omschrijving van de status quo waarbij hij zich aansluit.

De grote aandacht die selectieve toetsing zowel van studenten als docenten opeist, resulteert in een overmatige aandacht voor de eenvoudige kennis, en toepassingsproblemen, terwijl de veel relevantere inzichtelijke verwerking van de stof, het creatief en kritisch kunnen omgaan met materiaal uit het eigen wetenschapsgebied, te weinig aan bod kunnen komen.

Dit nadelig effect is zeker niet inherent aan iedere vorm van toetsing van leerresultaten, maar zal bij gebruik van toetsings-resultaten voor het nemen van beslissingen met selectief karakter (door docenten of door de student zelf) dit effect meestal in hoge mate hebben. Dat beoordelingsmethoden een actief bij het onderwijs proces in te schakelen hulpmiddel kunnen zijn, zonder daarbij met cijfers of zak/slaag beslissingen te hoeven werken, is een zienswijze die pas aarzelend zijn intrede doet

(=eigenlijk

43

herontdekt wordt) (Bloom, Hastings en Madaus 1971 bijvoorbeeld, voor een zeer uitvoerige bespreking en illustratie met ervaringsgegevens).

Veel onderzoek naar de beschreven samenhangen is er nog niet gedaan. Zo zegt McKeachie (1963) in een overzicht over het bereikte in het Amerikaanse onderzoek van hoger onderwijs: *"Gezien het belang van beoordelingen voor zowel studenten als docenten, is het te betreuren dat er zo weinig empirisch onderzoek over beschikbaar is. Hoe leren studenten hun eigen vorderingen te beoordelen? Hoe leren ze zichzelf bepaalde doelen te stellen? Hebben verschillende beoordelingsmethoden verschillende invloed op dergelijke leerprocessen? Kunnen er niet meer onderwijskundige alternatieven voor het cijfersysteem worden ontworpen?"*

Selectie in de loop van het onderwijsproces impliceert vrijwel onontkoombaar een competitief gedrag van studenten. Ieder voor zich, en God voor ons allen. Hofstee (1970, Universiteit en Hogeschool) heeft de competitieve momenten van "klassieke" beoordelingsvormen beschreven, o.a.: *"In gevallen waarin individuele beoordelingen*

worden gebruikt om "af te romen", d.w.z. om door vergelijkende examens een gefixeerd aantal plaatsen op te vullen met diegenen die de beste prestaties leveren, is het competitieve karakter het duidelijkst en wordt het door de betrokkenen ook doorgaans wel waargenomen (tenzij natuurlijk de hele beslissings-procedure zich aan hun waarneming onttrekt). Een goede prestatie van de één gaat hier automatisch ten koste van de slaagkans van de anderen", en: "Veel vaker wordt een systeem gehanteerd dat anders werkt, hoewel het ongeveer op hetzelfde neerkomt: men gaat uit van de gemiddelde prestatie, en kent een voldoende toe aan al degenen die daar niet te ver onder zitten. Dit systeem nu is echter nauwelijks minder competitief dan het "afromingsstelsel". 'Immers, iedere goede prestatie verhoogt ceteris paribus het gemiddelde en verhoogt dus de slaagkans van de anderen, wier prestaties met het gemiddelde worden vergeleken".

Over het verband tussen dergelijke vormen van beoordeling en bovenstaand gedrag van studenten zegt Hofstee: *"Redeneringen als de bovensaande zijn vrij subtiel en ik wil in de laatste plaats*

- 44

beweren, dat zulke implicaties voor iedere tentaminandus duidelijk zijn, wel moet gezegd

worden, dat ze strikt logisch uit de geschetste beoordelingspraktijk voortvloeien. Belangrijker is nog dat, zoals de leerpsychologie overtuigend heeft aangetoond, mensen zich naar zulke implicaties kunnen gaan gedragen ook zonder zich ervan bewust te zijn. Er is dus reden om aan te nemen, dat de boven beschreven beoordelingsstelsels het concurrentiemotief aanwakkeren, en dus strijdig zijn met het bevorderen van een koöperatieve instelling; temeer omdat ze harmonisch aansluiten op andere competitieve beloningsstelsels in onze maatschappij.

Maar niet alleen de relaties tussen studenten onderling worden op deze wijze in een ongewenste richting gestuurd. Ook contacten tussen student en docent worden door de ongelijke relatie die altijd tussen beoordelaars en beoordeelden zal bestaan, vertroebeld. Het resultaat is veelal een gebrek aan vertrouwen wederzijds en een tekort aan kommunikatie dat gaat ten koste van verbeteringen in het onderwijs en een goede begeleiding van de student. Dat tekort aan kommunikatie blijkt onder andere uit het verschijnsel dat veel docenten niet op de hoogte zijn van de aantallen studenten die voor het tentamen waarvoor zij verantwoordelijk zijn, afgaan of zelfs niet op hebben durven komen. Dat

dergelijke informatie "verdrongen" wordt is een rem op de ontwikkeling naar nieuwe vormen van onderwijs in overeenstemming met veranderde inzichten. [is dit inderdaadog steeds geciteerd uit Hofstee?]

De nadelige effecten die selectie heeft op de kwaliteit van het onderwijs en daarmee samenhangende zaken zijn moeilijk te becijferen. Dat neemt niet weg dat hier één van de belangrijkste punten van overweging ligt bij het besluit om al dan niet te selekteren tijdens het onderwijs. Ook kosten die moeilijk te berekenen zijn moeten afgerekend worden, al zal de afweging van voor- en nadelen van selectie er misschien niet makkelijker op worden. Waar het echter zó is dat andere kostenfactoren het potentiële rendement van selectie al vrijwel "opgegeten" hebben, zoals beargumenteerd in de eerste hoofdstukken, lijkt het redelijk om onderwijskundige argumenten in dat geval de doorslag te laten geven. Dan is selectie tijdens het onderwijs een verliesgevende strategie.

45

ANDERE PROCEDURES

Wanneer selectie tijdens het onderwijs af te raden is, welke mogelijkheden resten ons dan?

Selektie vooraf zal voor een deel stuiten op dezelfde moeilijkheden en bezwaren als in dit rapport genoemd ten aanzien van selektie door middel van propedeutische examens e.d.. Selektie vooraf betekent daarentegen wél dat in het volgende onderwijs alle tijd en energie ook inderdaad aan onderwijstaken kan worden besteed, en dat meer koöperatieve vormen van onderwijs tot de reële mogelijkheden gaan behoren. De kosten van de selektie-procedure zèlf zullen echter in dezelfde orde van grootte liggen als bij het propedeutisch examen. Bovendien moet rekening gehouden worden met onvermijdelijke persoonlijke onrechtvaardigheden die aan iedere selektie-procedure op basis van prestatie-toetsen inherent zijn.

Hoewel selektie vooraf onmiskenbare voordelen heeft boven systemen waarin tijdens het onderwijs geselekteerd wordt, is het een kostbare procedure en kunnen tegen de gang van zaken principiële bezwaren gemaakt worden, zoals deze op verspreide plaatsen in dit rapport naar voren gebracht zijn.

Dan is er nog een mogelijkheid van geheel andere orde, al van ouds bekend als selektie middel voor een ander netelig probleem waarin kwesties van sociale rechtvaardigheid een

grote rol spelen, namelijk het oproepen van dienstplichtigen voor het leger. Het klinkt belachelijk om in alle ernst voor te stellen te gaan loten voor een plaatsje in het hoger onderwijs. Maar laten we ons zelf niet voor de gek houden, immers een selectie-procedure op basis van gegevens met een slechts geringe voorspellende geldigheid komt goeddeels op hetzelfde neer. En ook wanneer de voorspellende geldigheid groot is, kunnen er bezwaren tegen geuit worden door personen of groepen die zich systematisch ten achter gesteld voelen in de gehanteerde examen-methoden. Het valt mij zwaar andere modellen van lotings-procedures te bedenken dan de juist genoemde schijn van belachelijkheid die aan het voorstel kleeft. De lijst van voordelen is echter erg lang, en misschien vinden velen hun bezwaren in de hier gepresenteerde lijst van voordeelen terug. Te bedenken is dan dat de bespreking van de lotingsprocedure zeer bewust uitgesteld is

- 46 -

tot het allerlaatst, omdat voor een nuchtere afweging van de voor en nadelen inderdaad al het voorgaande nauwelijk gemist kan worden.

Daar gaan we dan:

Loten is op zich een goedkope procedure, het kost de maatschpppij geen geld en de student geen

tijd.

Loten heeft geen negatieve invloed op voorafgaand onderwijs, niemand kan zich er op voorbereiden (tenzij men op bepaalde wijze gekwalificeerd moet om aan de loting mee te mogen doen, dit is echter geen problematische kant van de lotingsprocedure want de moeilijkheden liggen in dat geval in de selectieve situaties in voorafgaand onderwijs).

Loten is een zeer onthullende aangelegenheid, en wel het meest voor degenen die uitgeloot worden.

De politieke beslissing die ten grondslag ligt aan de noodzaak slechts een deel van de gegadigden tot het onderwijs toe te kunnen laten, wordt niet ingekleed in allerlei versluierende procedures. De slachtoffers weten waar ze moeten protesteren (en dat ze kunnen protesteren, een examenresultaat daarentegen heeft iets "onaantastbaars" (ten onrechte overigens, zoals we hebben gezien).

Uiteraard is loten een procedure die aan het onderwijs voorafgaat, het is niet zinvol haar pas na enige tijd - een jaar bijvoorbeeld - te houden. Selectie door loting impliceert dan ook een daaropvolgend onderwijs dat in belangrijke mate selectievrij kan zijn.

Bij loten kan niet het bezwaar gemaakt worden dat ongelijke monniken gelijke kappen opgezet

worden: dat zou het geval zijn bij selectieve toetsing. Bij selectie door toetsing hebben we te maken met voor iedereen dezelfde procedure, waaruit gelijktijd ongelijke kansen volgen, omdat er vrij grote persoonlijke verschillen zijn in mate en aard van voorbereiding, specifieke belangstellingen en begaafdheden, ingespeeld zijn op de toetsingssituatie zelf, etc. Bij loting zijn de kansen strikt gelijk.

Degenen die door loting afgewezen worden, krijgen niet het odium op zich geladen ergens voor "mislukt" te zijn, een kneusje te zijn. Hun kansen om elders aan de slag te komen of onderwijs te krijgen zijn niet ongunstig beïnvloed door slechte examenresultaten.

De kosten en opbrengsten van een lotingsprocedure moeten vergeleken worden met de kosten en opbrengsten van andere mogelijke vormen van selectie. Het gehele voorgaande betoog heeft geresulteerd in de constatering dat van selectieve examens al met al geen gunstig rendement kan worden verwacht, dat de er aan verbonden kosten - financieel, mankracht, verlies van kwaliteit v.h. onderwijs, persoonlijk welzijn v.d. student - dat deze kosten hoog, erg hoog zijn. Loten verdient dan een zekere voorkeur.

LITERATUURLIJST.

- Begeer, W.: "Numeriek Rendement", het selectieproces in het wetenschappelijk onderwijs, Wolters-Noordhoff, Groningen, 1968.
- Bloom B.S.; Hastings, J.T.; en Madaus, G.F.: "Handbook on formative and summative Evaluation of Student Learning", McGraw Hill, New York, 1971.
- Carroll, J.B.: "A model of school learning", Teachers College Record, 1963, 64, 723-733.
- Cronbach, L.J.: "Test Validation", in: Thorndike (Ed.) "Educational Measurement", 1971.
- Cook, w.w., The functions of measurement in the facilitation of learning In Lindquist (Ed.) "Educational Measurement", 1951.
- Cronbach, L.J.: "Essentials of Psychological Testing", 3e editie, Harper & Row, New York, 1970.
- Cronbach, L.J. en Gleser, G.C.: "Psychological tests and personnel Decisions, University of Illinois Press, Chicago, 1965.
- Doornbos, K.: "Opstaan tegen het zittenblijven", Stichting voor Onderzoek van het Onderwijs, Staatsuitgeverij, 1970.
- Fishman, "Some Social Psychological Theory for Selecting and Guiding College Students", in N.Sandford (Ed.) "The Americian College", 1962.

- De Groot, A.D.: "Methodologie" Mouton, Den Haag, 1961
- De Groot, A.D.: "Vijven en Zessen, Cijfers en beslissingen: het selectieproces in ons onderwijs". Wolters-Noordhoff, Groningen, 1966.
- De Groot, A.D.; Van Naerssen, R.F., e.a.: "Studietoetsen", Mouton, Den Haag, 1969.
- Halsey, A.H., Floud, J., en Anderson, C.A. (eds.): "Education, Economy, and Society", Free Press, New York, 1961.
- Hazewinkel, A.: "Selectie van studenten", publicatie van afdeling onderwijs research, Medische faculteit Rotterdam, ook verschenen als artikel in Intermediair, 29 oktober 1971.
- Heek, F. van, en anderen: "Het verborgen talent", milieu, schoolkeuze en schoolgeschiktheid. Boom, Meppel, 1968.
- Hofstee, W.K.B.: "Beoordelingen van studieprestaties", Universiteit en Hogeschool, november 1970.
- Hofstee, W.K.B.: "Selectie van personen", inaugurele rede, 1970.
- Humphreys, L.G.: "The fleeting nature of the prediction of College and academic success". Journ.of Educational Psychology, 1968, 59, blz. 375 e.v.
- Lavin, D.E.: "The prediction of academic performance", Wiley, New York, 1965.

- Lindquist, E.F. (Editor): "Educational Measurement", American Council on Education, Washington, 1951.
- Lord, F.M., en Novick, M.R.: "Statistical Theories of Mental test Scores", Addison-Wesley, London, 1968.
- Meer, A.J.van der: "Tussen Start en Finish", een onderzoek naar het studieverloop van de studenten Sociologie en Psychologie aan de V.U., 1955-1969. Sociaal Wetenschappelijk Instituut. V.U., Amsterdam.
- Meuwese, W.; Nelissen, Mevr.N., Tielens, H. "Een onderzoek naar de samenhang van enkele factoren met studieduur", T.H.Eindhoven 1969
- Neerzien, R.F. van, : "Het nut van selectie, enkele voorbeelden", in: Drenth, Willems en de Wolff (red.): "Bedrijfspsychologie", Kluwer, 1970.
- Onderwijs en Studiebegeleiding in het eerste jaar. Rapport van de kommissie bestudering konsekventies nota studiebegeleiding Rijksuniversiteit Utrecht, 1969.
- Ontwerp van Wet (en memorie van toelichting) herstructurering wetenschappelijk onderwijs; vrijwel gelijk aan het voorontwerp, Staatsuitgeverij, 's Gravenhage, 1971.
- Posthumus, K.: "De universiteit ~ doelstellingen, functies, structuren, oktober 1968.
- Idem eerste voortgangsnota, december 1969.
- Idem, tweede voortgangsnota, april 1970.

- Staatsuitgeverij, 's Gravenhage.
Rapport Studiegroep Selectie, T.H.Twente,
Enschede 1971.
- Rapport van de commissie studieduur, publicatie
van de Academische Raad, november 1969.
- Rulon, P.J.:; Tiedeman, O.V.; Tatsuoka, M.M., en
Langmuir, C.R.: "Multivariate Statistics for
Personnel Classification", Wiley,, New York,
1967.
- Sanfords, N. (Ed): "The American College", Wiley,
New York, 1962.
- Schwarz, P.A.: "Prediction Instruments for
Educational Outcomes", in: Thorndike (Ed.):
"Educational Measurement", 1971.
- Stanley, J.C.: "Reliability", in: Thorndike
(Ed.): "Educational Measurements", 1971.
- Thorndike, R.L.: "Personnel Selection, Test and
Measurement techniques", Wiley, New York,
1949.
- Thorndike, R.L.(Ed.): "Educational Measurement",
second edition, American Council on
Education, Washington, 1971.
- Wijnen, W.H.F.W.: "Onder of boven de maat", een
methode voor het bepalen van de grens
voldoende/onvoldoende by studietoetsen,
Groningen, 1971.
- Willems, P.J.: "Voorspelbaarheid van
studiegeschiktheid voor hoger onderwijs".
Nijmegen, 1959.

Willems, P.J.: "Een proefschrift beproefd",
Ned.Tijdschrift v de Psychologie, 1964, XIX,
blz.124-136.