

[Persoonlijke aantekeningen bij het destijds geruchtmakende boek waaraan Adriaan de Groot de belangrijkste bijdrage leverde (Arie van Peet schreef bijlage over onderwijsstelsels elders): Selektie voor en in het hoger onderwijs. COHO. Dit is een nogal ruwe en breedsprakige tekst, die bovendien lijdt aan het euvel van niet aan elkaar geschreven woorden (in zowel teksten van De Groot, als in mijn aantekeningen). Afijn, het beperkte belang van deze aantekeningen is dat ze een oefening zijn geweest in de aanvang van een halve eeuw stukken en publicaties over selektie etcetera (ook intelligentie, ha)]

Hoofdstuk 2 de noodzaak van selectie.

"Een consequentie van deze behoefte aan probleemoplossers is, dat het "intellectualisme" en de 'meritocratische' tendenties die onze maatschappijvorm vaak wordt verweten, voor een groot deel barre noodzaak zijn,. Wij moeten wel de moeilijkste problemen en daarmee een grote invloed aan de, naar wij hopen, bekwaamsten (deskundigen en bestuurders) delegeren."(15)

"bekwaamheid is schaars" (16) je vraagt je bij dit soort uitspraken wel af of ze niet volkomen triviaal zijn. Tegenwoordig is zuivere lucht en schoon water ook schaars. Voor een bepaald soort arbeid in onze maatschappij trekken we mensen uit het middellandse zee gebied aan. Anderzijds wordt schaarste zeker ook bepaald door de vraag die er naar het betreffende artikel is, en in het licht van de toenemende werkloosheid onder academici vraag je je af of bepaalde soorten hoe bekwaamheid wel zo schaars zijn als De Groot wil doen voorkomen. Zoveel is zeker, dat wil deze uitspraak van de Groot enige betekenis hebben, dan zal hij gekwantificeerd moeten worden. Zolang dat niet gedaan is is het nogal voorbarig om op dit soort vaagheden conclusies te gaan baseren zoals in de loop van dit rapport gebeurt. (schermen met het "Peter principle" om een en ander aannemelijk te maken is natuurlijk te gek om ongestraft voorbij te laten gaan)

Volslagen irrelevant is een vergelijking tussen het aantal zwakzinnigen en en het aantal intellectuelen in een land. (blz 16 e.v.)

"Uit eigen schoolervaring en, meer nog, uit de ervaring van onderwijzers en leraren (m.n. in het l.b.o.) weten wij, dat begrip voor zelfs elementaire wiskunde maar aan weinigen gegeven is; de grote meerderheid mist zowel het nodige leervermogen als de vereist intellectuele belangstelling."(16)

Wat moeten we met dergelijke tendentieuze uitspraken aan?

"bijna niemand echter kan hogere wiskunde leren, schaakmeester, pianist, redenaar, wetenschapsbeoefenaar worden." (17)

Het percentage van de Nederlandse beroepsbevolking dat valt in beroepscategorieën als door De Groot hier genoemd, zal aanzienlijk zijn. Vergelijk ook deze uitspraak met het volgende citaat uit het artikel van Jensen waar ook De Groot later naar verwijst:

"Some minimal level of ability is required for learning most skills. But while you can teach almost anyone to play chess, or the piano, or to conduct an orchestra, or to write prose, you cannot teach everyone to be a Capablanca, a Paderewski, a Toscanini, or a Bernard Shaw."

"Gebleken is dat niet alleen het school en carrière-succes van kinderen maar ook hun intelligentie (ongeacht met welke test gemeten) sterk met het intellectueel en beroepsniveau (en daarmee de sociale klasse) van de ouders samenhangt." (19)

Dit is iets dat in ieder handboek op het gebied van intelligentie onderzoek met behulp van tests te vinden moet zijn, volgens de Groot. Welnu: Jensen: (HER XXXIX blz 478) "An estimate of the average correlation between occupational status and IQ is .50." Jensen (HER XXXIX blz 15,16) citeert: onderzoek van Duncan

"Duncan's (1968) detailed analysis of the nature ^{xx} of the relationship between intelligence and occupational status led him to the conclusion that 'the bulk of the influence of intelligence on occupation is indirect, via education.' If the correlation of intelligence with education and of education with occupation is, in effect, 'partialled out,' the remaining 'direct' correlation between intelligence and occupation is almost negligible." Eenzelfde analyse levert ook de conclusie: "men with the same schooling and in the same line of work are differentially rewarded in terms of mental ability". Jensen (zelfde artikel. blz 58): "The heritability of measures of scholastic achievement is much less, on the average, than the heritability of intelligence. In reviewing all the twin studies in the literature containing relevant data, I concluded that individual differences in scholastic performance are determined less than half as much by heredity as are differences in intelligence.

Jensen (zelfde artikel blz 75) "It is well known that children's IQ's, by

school age, are correlated with the socioeconomic status of their parents, This is a worldwide phenomenon and has an extensive research literature going back 70 years. Half of all the correlations between social class and children's IQ's reported in the literature fall between .25 and .50, with most falling in the region of .35 to .40. Cattell (abilities, 1971) over verband tussen intelligentie en schoolprestaties: (blz 378) "In general, correlations of intelligence test with school performance that exceed .6 or even .5, depending on the population) show that the intelligence test is itself contaminated with school achievement and is anything but a pure culturefair measure."

De correlatie tussen intelligentie van vaders met hun kinderen is wat zeer laag is vergeleken met de .9 genotype phenotype cor3 die blijkt uit de gevonden nature nurture verhouding van .8 ong,

Hiernaast kunnen nog gegevens over sociale mobiliteit gezet worden, die zeker in Amerika en w. en n.w. Europa zeer hoog is.

Een en ander leidt tot de conclusie dat de apodictische uitspraak van De Groot zoals aangehaald bovenaan blz 3 verzwakt moet worden van een "sterke samenhang" naar "een samenhang". Omdat dergelijke verzwakkingen van zijn uitspraken herhaaldelijk gemaakt moeten worden, vraag je je af of er nog iets anders dan holle fraseologie overblijft van deze inleidende hoofdstukken. Voorzover de voorstellen die concreet door De Groot gedaan worden, hun oorsprong vinden in dit soort zwakke beschouwingen, worden ook deze voorstellen twijfelachtiger naarmate zijn uitgangspunten ontkracht blijken te kunnen worden.

Ook de andere resultaten van intelligentieonderzoek, zoals die door de Groot gezien worden, zijn ofwel triviaal dat zich tussen de groepen der moeilijk lerenden en der intellectueel begaafden, dus binnen de grote midengroep der "normalen", nog zeer grote verschillen in intelligentie (leervermogen) voordoen") ofwel idiosyncratisch ("dat de intelligentie van de in dit opzicht "gemiddelde Nederlander" ontoereikend is om de in 2.2 en 2.3. bedoelde problemen in hun complexiteit te kunnen begrijpen, laat staan op te lossen")

2.5 erfelijkheid en milieu.

"Het algemene verband is dit: hoe hoger ontwikkeld het onderwijs systeem, hoe beter de detectie en doorstroming van intellect in het onderwijs proces, des te hoger is het erfelijkheids aandeel in de

intelligentie variantie en des te geringer is de resterende reserve."

Het eerste deel van deze uitspraak is ongetwijfeld juist, en wordt nog beter benadrukt door Carl Bereiter (HER XXXIX blz 3101);

"The heritability of intelligence is unquestionably high, but what is more to the point is that with further social progress its heritability can only increase, because of the elimination of such sources of environmental variance as differences in the quality of education, nutrition, and medical care. One's view of the future beyond equality of opportunity must, therefore, be of a future in which differences in intelligence are virtually one hundred percent determined by heredity."

Waaraan hij in een voetnoot toe voegt:

"This eventuality is in no wise to be forestalled by individualized instruction or any more libertarian tactic; on the contrary, such approaches should allow inherited differences to reach full flower, as advertised in the slogan, "enabling each child to realize his fullest potential."

De merkwaardige uitspraak van De Groot dat hoe hoger het erfelijkheids aandeel in de intelligentie variantie, des te geringer de resterende reserve, kan toevallig gecontrasteerd worden met de zin die volgt op het Juist gegeven citaat van Bereiter:

"The magnitude of these differences then becomes a crucial question; however, it is not magnitude in terms of IQ points that counts, but magnitude in terms of differences in effective problemsolving capability. These are not the same thing, even if they are perfectly correlated. "We may expect that through continued scientific progress, through the continued development of intellectual "tools", such as language, logic, thinking machines, and scientific techniques, man's ability to solve problems will increase at an accelerating pace, even though his IQ changes not a jot. There is nothing paradoxical in this. Through the development of mechanical tools, men's ability to lift weights, hurl objects, and so on, has been multiplied manifold, independently of any increase in his basic muscular strength."

M.a.w. blijkt het heel eenvoudig mogelijk een even plausibele redenering als die van De Groot op te zetten, die echter op heel andere implicaties schijnt te wijzen.

In de discussie die nu al is losgebarsten over met name de uitspraken van De Groot over de erfelijkheid van intelligentieverschillen, lopen een

aantal zaken door elkaar. Ik zal proberen er een aantal van op een rijtje te zetten.

! misvattingen over het begrip "intelligentie" met daaraan gekoppeld een simpele afwijzing van de door de Groot genoemde verbanden.

! gebrek aan inzicht in de strekking van de uitspraak dat 80 % valide variantie in intelligentie toegeschreven moet worden aan erfelijke factoren.

! met name niet op de hoogte zijn van het verschijnsel dat hoe gelijkjer de kansen voor iedereen op goed onderwijs, hoe meer erfelijk bepaalde verschillen ook de phenotypische verschillen zullen gaan bepalen.

! met name ook een door elkaar halen van verband tussen genotype en phenotype, en de correlatie tussen intelligentie van ouders en hun kinderen.

! met name ook het verwarren van uitspraken die uitdrukkelijk alleen geldig zijn voor grote populaties, met uitspraken over bepaalde individuen.

! overschatting (zeker ook door De Groot) van de rol van intelligentie als bepaler van school en maatschappelijk succes

! de, enigszins ridicule, misvatting dat uit hoge erfelijkheid etc geconcludeerd mag worden dat het geven van onderwijs dan een grotendeels overbodige bezigheid zou zijn

! de voorbarige conclusie dat iedere vorm van remedial teaching, en van compensatorisch onderwijs, gedoemd zou zijn tot mislukken.

! de idee dat intelligentie iets is dat in de eerste levensjaren thuis, en later vooral op school 'geleerd' wordt, en dat het dus niet erfelijk bepaald kan zijn.

! mijns inziens ook een misvatting van velen is, dat de maatschappij ideaal zou zijn, wanneer we een vorm van onderwijs hebben waarin iedereen gelijke kansen heeft om zijn eigen potenties ten volle te realiseren. Een dergelijke primitieve opvatting over een 'socialistische' maatschappij is inderdaad volstrekt in strijd met onderzoekgegevens als de door De Groot genoemde. Een mooie gelegenheid om er nog eens op te wijzen dat gelijkheidsidealen bereikt moeten worden door maatschappijveranderingen, en niet door versiering van het onderwijssysteem. Ook Chomsky analyseert de problematiek vanuit deze invalshoek (Ramparts juni).

! aantijgingen van bevooroordeeldheid van De Groot en van de onderzoekers waarop hij zich baseert, moeten natuurlijk niet verward worden met een behoorlijke discussie over de draagwijdte van dit soort onderzoek gegevens, bijv. m.b.t. implicaties voor het onderwijs. (dat neemt niet weg dat De Groot weinig moeite heeft gedaan om

misverstanden zoals die zojuist opgesomd zijn, te voorkomen).

2.6 het intellectueel potentieel van het Nederlandse volk.

Ook in deze paragraaf weer een opeenstapeling van tendentieuze, op zijn minst niet van een behoorlijke argumentatie voorziene, uitspraken "Kortom: in een qua onderwijs hoog ontwikkeld en qua sociale structuur betrekkelijk homogeen land als Nederland moet men aannemen: (1) dat het intellectueel potentieel, d.i. de hypothetische verdeling van de intelligentie aanleg, er niet veel beter zou uitzien dan de tot bescheidenheid (?) nopende verdeling van de effectieve (?) intelligentie, (2) dat er nog maar relatief weinig te verwachten is van aan te boren intellectuele reserves (?), (3) dat met name het probleem van de schaarste van bekwame probleemoplossers, ondanks de toenemende instroom in het hoger onderwijs, niet minder nijpend (?) zal worden dan het al is, en (4) dat de "barre noodzaak" (?) van selectie en goede opleiding van de bekwaamsten, gegeven de toenemende behoefte (?), alleen maar sterker kan (?) worden." (21a, 21b)

Hfdst 3

Het expeditiemodel van De Groot is voor het Nederlandse tertiaire onderwijs een nogal revolutionair idee. Voorbeelden van het functioneren van dergelijke onderwijsvormen zijn makkelijk te vinden. Of deze voorbeelden zonder meer overgezet kunnen worden naar de Nederlandse situatie mag op een aantal gronden betwijfeld worden. Daartegenover staat dat het huidige model van selectie onderweg, het veldloop model, dermate grote bezwaren, van zeer uiteenlopende aard, met zich meebrengt dat vrijwel ieder ander model dat deze voortdurende selectie tegengaat een verbetering van het onderwijs zal opleveren.

Dat betekent dat kritiek op het expeditiemodel van De Groot gezien moet worden als een stap op weg naar verbetering van dit zelfde model, en zeker niet als leidend tot afwijzing van het model ten gunste van de huidige vorm van universitair onderwijs.

Het is jammer dat De Groot ook bij dit onderwerp al te spaarzaam is in het noemen van bronnen. Aangenomen mag worden dat zijn expeditiemodel niet alleen gebaseerd is op de waarneming dat er onderwijsvormen functioneren waarin tijdens het onderwijs nauwelijks enige selectie meer plaats vindt, onderwijsvormen die ook in Nederland

eens uitgeprobeerd zouden kunnen worden. Er loopt daarnaast een duidelijke parallel tussen het expeditiemodel en een onderwijsmodel dat ook in Nederland steeds meer in discussie (en praktijk) komt, geassocieerd met onderzoekers als Bloom en Carroll en beter bekend onder de naam: "learning for mastery". Belangrijke elementen uit deze benadering, en de er achter liggende filosofie, zijn herkenbaar in het expeditiemodel.

"de ingangselectie voor selectievrij onderwijs is een selectiedrempel: wie deze gepasseerd is, is als op te leiden en te diplomeren student aanvaard."(28)

De Groot kiest voor een onderwijsmodel waarin onderweg niet meer geselecteerd mag worden, een keuze die op zeer goede gronden verdedigbaar is. Een dergelijk onderwijsmodel draagt echter als implicatie met zich mee dat er een ingangsdrempel moet bestaan die het mogelijk maakt dat iedere aangenomen student binnen bepaalde tijd tot een bepaald niveau opgeleid kan worden.

Op zich is een ingangselectie niet iets om van terug te schrikken, in het veldloop model vindt ook selectie plaats, maar dan tijdens het onderwijs, waarbij de nadelen, niet alleen voor de betrokken studenten, maar ook voor de maatschappij, aanzienlijk zijn. Dat neemt echter niet weg dat een dergelijke ingangselectie bij het expeditiemodel toch een aantal problematische aspecten in zich draagt, waar De Groot aan voorbij gaat, of waar hij niet zo zwaar aan tilt.

Bij het veldloop model is het zo, dat de mensen die verdwijnen het inderdaad niet meer aan kunnen, of het idee hebben het niet meer aan te kunnen. (andere motieven om met de studie op te houden gelden evenzeer voor het expeditiemodel en kunnen dus buiten beschouwing blijven). Natuurlijk is het altijd mogelijk dat een verbetering van het onderwijs, of wat meer aandacht voor de specifieke problemen van deze studenten, hun voortijdig verdwijnen had kunnen voorkomen; ook dit echter is een redenering die evengoed voor het expeditiemodel geldt. Het probleem bij het expeditiemodel wordt nu, om van te voren de mensen aan te wijzen, die de studie zoals die onder het expeditiemodel georganiseerd is, niet aan zullen kunnen (of, wat. in dit geval hetzelfde is, waarvan het niet mogelijk blijkt om ze met de nodige extra inspanningen van docentenzijde te brengen tot het vereiste minimum eindniveau). De Groot gaat in het rapport enkele malen op dit soort selectieproblematiek in, echter niet in dit verband, dus vanuit de vraagstelling in hoeverre het

mogelijk is voor een expeditiemodel een voorselectie op te zetten die aan een aantal belangrijke, aan dergelijke selectie te stellen eisen voldoet. De Groot gaat wel uitgebreid in op de vraag hoe een dergelijke selectie in zijn werk zou moeten gaan, hoe hij opgezet moet worden, welke waarborgen geschapen moeten worden, waar wel en waar niet naar gevraagd mag worden etc. etc" maar maar hij zich niet over uit laat is over de van een dergelijke procedure te verwachten kwaliteit, met name het aantal foute beslissingen dat onder een aantal plausibele aannamen bij dergelijke selectieprocedures gemaakt zal worden. Nogmaals, een dergelijke analyse hanteert hij wel op een aantal andere plaatsen in hetzelfde rapport, maar heeft hij niet toegepast op de onderhavige vraagstelling. Dit is daarom zo belangrijk, omdat op grond van een aantal overwegingen verwacht mag worden dat bij dergelijke selectieprocedures een niet te verwaarlozen percentage van de afwijzingsbeslissingen, maar ook van de toelatingsbeslissingen, foutief zal zijn. Het expeditiemodel kan wél zo ingericht worden dat niet al te ernstige fouten van het type: foutief geaccepteerd, binnen het model van adaptief onderwijs opgevangen kan worden, maar het jammere van selectieprocedures is dat het zo klein mogelijk maken van het aantal van dergelijke foutieve toelatingsbeslissingen resulteert in het groter worden van de aantallen foutieve afwijzingsbeslissingen.

Een ander punt waaraan de nodige aandacht nog besteed moet worden, is de nadruk die er bij De Groot valt op de voorwaarde die de faculteit stelt voor toelating: nl. dat de faculteit de zekerheid moet hebben dat zij de betreffende student met behulp van de ter beschikking staande onderwijsmiddelen en mankracht binnen de daarvoor gestelde tijd tot het gestelde minimum niveau vereist voor goede afsluiting van de cursus, kan brengen. Dit werkt, zoals de Groot het formuleert, een zeer conservatieve toelatingspolitiek in de hand. Tenminste moet een dergelijke regeling dan gewijzigd worden in die zin, dat ook de faculteit gedwongen wordt bij dergelijke beslissingen een bepaald eigen risico te nemen.

Een voorbeeld van een examen dat uitstekend zou kunnen passen in het expeditiemodel, is de praktijk aan het Medical College of Virginia, waar 2 of 3 van de honderd studenten voor een examen plegen te zakken. Dit is een volkomen willekeurig voorbeeld, dergelijke procedures komen bijv. in het Amerikaanse hoger onderwijs veelvuldig voor. (het genoemde voorbeeld is te vinden in The Journal of Experimental Education, vol 39, 1971, blz 61-63. S.J.Kilpatrick: an alternative to the standardized score in grading a multiplechoice examination.

Nog even terugkomend op de voorlaatste alinea. De Groot formuleert aldus:

"Dit is een kwestie van afstemming; namelijk afstemming van aard en niveau van de selectie drempel op aard en niveau van het programma en het afsluitende examen. Die afstemming is zo geregeld, en deze operationele definitie is van fundamenteel belang, dat alle bij het programma betrokken docenten en examinatoren het erover eens zijn, dat zij, gegeven de vastgestelde minimum eisen van het programma, de opleiding van de aldus geselecteerden op zich kunnen nemen." en:

"De onderwijs capaciteit is berekend op de eisen van selectievrij onderwijs en wordt strikt gehandhaafd. Dit houdt in, dat er een numerus fixus is voor het per cursus op te nemen aantal nieuwe studenten. Deze numerus fixus is het maximum aantal, waarvoor de instelling gegeven de doelstellingen, de outillage, de aanwezige docenten, de hoofdlijnen van het onderwijsplan, inclusief de steunprogramma's en dgl. nog in staat en bereid is een selectievrij contract te "tekenen". Anders uitgedrukt: de numerus fixus is het maximum aantal nieuwe studenten, waarvoor de instelling het gecontracteerde studierendement van 100% kan garanderen." (34-35)

Dat De Groot voorbij gaat aan de selectieproblematiek voor het expeditiemodel blijkt bijvoorbeeld uit uitspraken als: "Geschiktheid: Onder het selectievrije contract is de geschiktheid van de student na zijn toelating geen vraagpunt meer. Die geschiktheid wordt, onder voorbehoud van de ijverclausule en de ziekteclausule, bij contract gegarandeerd, men kan ook zeggen: bij definitie vastgesteld." (37) Jawel, wie aangenomen is, is geschikt; maar op grond waarvan kunnen we dan iemand aannemen? Wanneer een selectie bedoeld is om de geschikten te selecteren, en we definiëren geschiktheid dan als het toegelaten zijn, dan zijn we met een cirkelredenering bezig.

Het enige wat er bij De Groot over de selectie gezegd wordt (vanuit de door mij geïntroduceerde probleemstelling):

"Concreter: Als het voorkomt, dat een goedwillende maar zwakke student er met een overmaat van "bijles" doorheen 'gesleept' moet worden en dan nog maar zozo aan de minimum eisen voldoet, dan schort er iets aan de selectiedrempel." (45)

De kritiek van De Groot op het traditionele psychometrische model is ten dele terecht, ten dele niet terecht. Zij is juist voorzover het het traditionele psychometrische model betreft. Zij is niet juist omdat het traditionele model niet noodzakelijk het psychometrische model is. De Groot heeft eerder in het rapport al verwezen naar het boek van Cronbach & Gleser waarin een summier schets van een groot aantal mogelijke psychometrische modellen gegeven wordt. Ook in Nederland wordt wel met dergelijke modellen, van de "traditie" afwijkend, gewerkt, o.a. Van Naerssen, coauteur samen met De Groot van 'het' boek over studietoetsen.

De Groot slikt een aanzienlijk deel van zijn kritiek dan ook weer in wanneer hij stelt:

"Het bovenstaande betekent niet, dat men door het uitvallen van studiesucces als bruikbare criterium variabele bij een selectievrije opzet geen middelen meer zou hebben om de voorafgaande selectie drempel op de traditionele wijze te valideren". (44) Maar ook deze opmerking neemt hij weer terug:

"Een traditionele "validatie" is dus wel mogelijk; maar, en dit is de crux van de redenering, zij heeft weinig betekenis."

Wat de Groot doet komt m.i. op het volgende neer: hij schetst een overdreven star beeld van de psychometrie, en laat het vervolgens voorkomen alsof toetsingsmomenten tijdens en voorafgaand aan het expeditiemodel zich niet meer zouden kunnen lenen voor welke psychometrische analyse dan ook. Een dergelijke troebele presentatie van zaken moet betreurd worden, komt in ieder geval een behoorlijke discussie, bijvoorbeeld over de toelatingsselectie, niet ten goede. Nog een keer. (Waar de Groot spreekt over de vooronderstellingen van de psychometrische werkwijze),

"Die vooronderstellingen zijn die van het veldloopmodel. Men zou ook van het junglemodel met de bijbehorende "survival of the fittest" kunnen spreken"(44) Het verwijt dat de Groot hier aan de psychometrie maakt, geldt veeleer liet onderwijs. Er is niets inherent aan de psychometrie dat de beschuldiging dat het werkt vanuit de vooronderstelling van het 'junglemodel' zou kunnen rechtvaardigen.

hfdst 4: examens: onderdelen, functies, soorten.

Het uit elkaar houden van de verschillende functies die vaak aan eenzelfde examen toegekend worden, is een goede zaak. Tegen één puntje moet echter bezwaar worden gemaakt nl.: "... dit houdt in dat "compensatie" van een tekort (onvoldoende) voor één onderdeel door een hoog cijfer voor een ander onderdeel uitgesloten is."

Dit is een aanbeveling van De Groot, waarvan invoering zou kunnen leiden tot een groot aantal onrechtvaardigheden. Het ware beter dat de auteur dergelijke aanbevelingen vergezeld deed gaan van een opsomming van eventuele voordelen, maar vooral van de mogelijke nadelen, en argumenten die tegen dergelijke procedures pleiten. (zie bijv. het artikel van Van Naerssen in het onderwijs research congresboek 1968) Zie ook R. L. Thorndike (Ed) Educational Measurement, blz. 694 e.v.

Het is jammer dat De Groot zonder meer uitgaat van een model waarin er selectie plaats vindt aan het einde van een propedeutisch jaar. Weliswaar doet hij een aantal voorstellen die enkele nadelen van een dergelijke selectie zouden moeten neutraliseren (bijv. door dit jaar een soort zelfstandige rol toe te delen, naast een op verder universitair onderwijs voorbereidende), maar een grondige analyse van een dergelijke opzet in vergelijking tot een vorm van onderwijs (en selectie) waarin eventuele selectie vóóraf plaats vindt (dus niet pas na een jaar) met handhaving van het expeditiemodel, maar dan voor het gehele onderwijs, zou interessanter zijn (en misschien tot een ander onderwijs beleid t.a.v het tertiair onderwijs kunnen leiden).

Wanneer je als gegeven accepteert dat de selectie na één jaar plaats vindt, dan zijn de ideeën van De Groot over de propedeuse zeer de moeite waard:

"Iedere propedeutische opleiding moet één uit de algemene doelstellingen vooropleiding, introductie en onderwijskundige samenhang afgeleid, maar zo gedifferentieerd mogelijk totaal programma annex afsluitend examen kennen, waarvoor ook een substantieel aantal van de toelatingsafgewezen studenten moet kunnen slagen." (59)

"Een belangrijke tweedeling is die in taken waarvoor objectieve, minimale prestatie-eisen kunnen worden gesteld (P), en andere taken waarvoor dit niet kan. Bij deze tweede categorie gaat het alleen om het verricht hebben van bepaalde, gecontroleerde en bijgeschaafde, handelingen

(H): een practicum gevolgd hebben, een voordracht gehouden hebben."
(blz. 51)

Commentaar: Een betere filosofie voor een dergelijk onderscheid lijkt mij het volgende; Onderwijsdoelstellingen zijn slechts gedeeltelijk te formuleren in termen van meetbaar eindgedrag. Dit meetbaar eindgedrag is ook maar gedeeltelijk geschikt om in de vorm van mondeling of schriftelijk examen beoordeeld te worden, Sommige doelstellingen kunnen geformuleerd worden in termen van het meegedaan hebben aan bepaalde activiteiten, het gemaakt hebben van werkstukken die voldeden aan bepaalde criteria e.d. Er blijven dan nog een groot aantal doelstellingen over die niet te vangen zijn onder P-eisen of H-handelingen. Voorbeelden van dergelijke doelstellingen zijn te vinden in de in de wet geformuleerde doelstellingen van het hoger onderwijs. Deze doelstellingen moeten veeleer gerealiseerd worden in termen van vormgeving, aard en inhoud van het totale onderwijsgebeuren. Dat betekent o.a. dat wat De Groot P en H onderdelen noemt, ondergeschikt moeten zijn aan meer algemene doelstellingen van onderwijs. En dát zou wel eens kunnen betekenen dat sommige P, maar misschien ook wel H onderdelen, in strijd zijn met andere doelstellingen.

Nog even voor de duidelijkheid: dit gaat niet over kwesties als acceptabiliteit en billijkheid etc. van een selectiebatterij die bestaat uit P en H onderdelen (daar zal De Groot in de volgende hoofdstukken uitgebreid aandacht aan schenken), maar over de verenigbaarheid van een en ander met andere doelstellingen van onderwijs (bijv. of een bepaald selectieprogramma wel verenigbaar is met een doelstelling die de ontwikkeling van maatschappelijk verantwoordelijkheidsbesef behelst, om maar een buitenplaats te noemen).

Een andere kritiek is, dat het P H onderscheid niet noodzakelijkerwijs verschillende onderwijs activiteiten impliceert.

Ook is het denkbaar, bijv, vanuit Bloom's masterylearning concept, dat de P onderdelen een zeer sterk H karakter gegeven kan worden. (De Groot geeft daar zelf een voorbeeld van, blz. 52 onderaan))

Tenslotte, in de toelichting hierop stelt De Groot dat de minima die voor de P onderdelen vereist worden, dan ook werkelijk objectief worden bepaald. Daartegen valt in te brengen dat een objectieve scoringsprocedure niet erg veel garanties geeft voor de objectiviteit van toets samenstelling, of van beslissingen die op grond van de scores

genomen worden. In deze zin is De Groot's pleidooi voor objectieve toetsing een beetje al te voorbarig. (zie ook blz. 15 van dit stuk)

4.2 afsluitende examens.

De Groot definieert het afsluitend examen als totaalexamen, d.i. het totaal van alle eisen waaraan een kandidaat moet hebben voldaan om te slagen (m.n. dus ook H onderdelen). Onderwijs en examen moeten elkaar wederzijds dekken.

De Groot laat zich niet uit over waaruit een goede 'dekking' kan blijken, Hij verwijst naar het psychologenjargon van inhouds- en begripsvaliditeit, Wat nu juist interessant is, is hoe De Groot zich voorstelt dat een behoorlijke validiteit in genoemde opzichten te bereiken is; ook welke eisen minimaal aan deze validiteit gesteld moeten worden; of het praktisch haalbaar is om schattingen van deze validiteiten te maken waarmee gewerkt kan worden. Hij zegt: "de verantwoordelijkheid voor deze dekking (.... berust bij de studieleiding daarmee in het midden latend hoe de studie leiding deze verantwoordelijkheid behoorlijk waar kan maken); en wat de consequenties zijn van (onvermijdelijke) tekorten in deze validiteitsopzichten. Het gevaar lijkt onontkoombaar dat het beste wat bereikbaar is, iets zal zijn in de trant van face validity. Dat is dan wel uitstekend verenigbaar met eisen van acceptabiliteit etc. maar betekent allerminst noodzakelijkerwijs een behoorlijke inhouds- of begripsvaliditeit. Te verwachten is dat de moeilijkheden bij bepalen van inhouds- en begripsvaliditeit even groot, zo niet groter zijn dan bij het bepalen van predictieve validiteit van examenprocedures.

Een ander betekent m.i. dat De Groot hier wat al te makkelijk een aantal eisen stelt waar geen enkele studieleiding (in voldoende mate) aan zal kunnen voldoen.

"Is die dekking verzekerd, dan kan men aan de hand van de (totale) examen resultaten nagaan of het gegeven onderwijs aan de doelstellingen heeft beantwoord."

Een goede dekking is erg moeilijk te verzekeren, zoals hierboven betoogd is. Maar ook al was dit wel het geval, dan belooft de geciteerde uitspraak te veel. Uit examenresultaten blijkt nooit zonder meer of het gegeven onderwijs aan de doelstellingen heeft beantwoord. Tenzij de examenprocedure en dataverwerking met erg veel (deskundige)

mankracht van te voren zo geprogrammeerd is dat die evaluatie er zonder meer uit rolt. Evaluatie van onderwijs in de hier door De Groot bedoelde zin, is een tak van onderwijsresearch die nog niet zo ver ontwikkeld is dat ze kant en klare recepten voor voor dergelijke evaluaties kan leveren (ook voor de toekomst mogen we geen hoop op dergelijke eenvoudige oplossingen hebben)

HOOFDSTUK 5: Acceptabiliteit van selectiemethoden.

5.1 acceptabiliteit als criterium.

"om acceptabel te zijn moet een selectieprocedure:(1) billijk zijn en (2) doelmatig, zijn." (72)

De moeilijkheid hiermee is, dat billijkheidseisen en doelmatigheidseisen met elkaar in strijd kunnen zijn (noot blz W Over de eisen van doelmatigheid van institutionele selectie (om nog maar helemaal niet te spreken van individuele selectie, waar De Groot in dit verband dan ook maar niet over spreekt) is erg weinig bekend (zie art. Hills in R.L. Thorndike (ed) Educational Measurement 1971), i.t.t. de suggestie van De Groot dat er veel bekend is: "Dat betekent niet dat er pasklare oplossingen, met selectiemiddelen en al, bestaan voor concrete selectieproblemen. Maar wel is bekend hoe men ze uit doelmatigheidsoogpunt in termen van betrouwbaarheid, validiteit, en selectiestrategieën voor een maximale opbrengst het beste kan aanpakken. Wat overblijft zijn 'technische' problemen." Waar De Groot deze, toch wel sensationele, uitspraken op baseert, maakt hij niet duidelijk, Over de doelmatigheid van zelfselectie (de beslissing van een bepaalde student om al dan niet (verder) te gaan studeren) is nagenoeg helemaal niets bekend.

"Eisen van billijkheid zijn eisen die naar eenstemmig oordeel, minimaal vervuld moeten zijn om te bereiken, dat een selectieprocedure, qua eerlijkheid van kansen voor de gegadigden, geaccepteerd kan worden."(72) en: "Billijkheid is: rationele verdedigbaarheid; een selectieprocedure is billijk als zij ontegenzeggelijk verdedigbaar is uit een oogpunt van eerlijke kansen (en verdedigbaar blijft bij een diepgaande kritische analyse, eventueel in een moeilijke overleg situatie)."(72)

Hoofdargumenten voor selectie zijn volgens De Groot:

"(1) Geschiktheid: Gegadigden van wie op goede gronden kan worden

verwacht, dat zij het onderwijs van het programma niet zullen kunnen volgen, worden niet toegelaten.

(2) Numerus fixus: Als het aantal gegadigden de op goede gronden vastgestelde capaciteit van een bepaald programma aan een bepaalde onderwijsinstelling overtreft, is afwijzing van een deel van de gegadigden, mits op goede gronden, onvermijdelijk.

De eigenlijke acceptabiliteitsproblemen liggen besloten in de vraag wat voor gronden goede gronden zijn."(74)

5.3 objectiviteit

"acceptabiliteit van een selectieprocedure is alleen bereikbaar als men iedere vorm van subjectieve beoordeling van geschiktheid uitsluit. Dit is bereikbaar, indien men uitsluitend onbetwistbare gedrags- en persoonsgegevens van de gegadigden verwerkt in een voor allen gelijke formule, die ook zelf als resultaat een beslissing (toelating of afwijzing) oplevert, òf een eindscore oplevert op basis waarvan men zo'n beslissing, in absolute of relatieve zin, kan nemen."

Dit uitgangspunt leidt binnen de kortste keren tot ernstige tegenspraken. Bovendien is het alternatief voor toelatingsprocedures waarin momenten van grote willekeur aanwezig zijn, niet een op de computer uitvoerbare beslissings strategie. De objectiviteit van De Groot is in niet onaanzienlijke mate een schijnobjectiviteit:

"Scores behaald op (objectief scorebare) toetsen en tests van allerlei aard voldoen wel aan de eis mits zij in een objectieve beslissingsformule worden verwerkt." Een weerlegging die minder triviaal is dan zij er op het eerste gezicht uit ziet, is dat de eerste de beste klungel die een meerkeuze toets in elkaar flanst, een instrument heeft gemaakt dat aan De Groot's objectiviteits eis voldoet. Ook De Groot weet wel beter, de vraag is alleen waarom hij in dit rapport van dat beter weten zo weinig laat merken. Objectieve scorebaarheid heeft met objectiviteit in andere opzichten bar weinig te maken. En juist dié objectiviteit levert grote theoretische en praktische moeilijkheden op, en zou bij een min of meer goede realisering wel eens in strijd kunnen komen met andere acceptabiliteits eisen. (Zie bijv: Bormuth: on the theory of achievement test items, University of Chicago, 1970)

5.4 differentiële rechtvaardigheid.

"De typische situatie, waarin onbillijkheid van genomen beslissingen kan

blijken, is die van de vergelijking van twee (grens)gevallen waarin tegen gestelde beslissingen zijn genomen(....) De stricte opvatting van acceptabiliteit vereist dat de examinerator zijn beslissing in zulke gevallen altijd kan verantwoorden." (76)

Aan dit topic wijdt De Groot een zware discussie, die m.i. enigszins bezijden de zaak waarom het gaat, staat. Wél belangrijk is bijv. een behoorlijke herkansings regeling. Daarover De Groot:

"Dat de mogelijkheid tot herkansing in het selectie systeem (m.n. bij de propedeutische selectie) ingebouwd méet zijn, is zo algemeen aanvaard, dat ook daarover niet veel behoeft te worden gezegd."(80)

En daar laat hij het eigenlijk bij. Over de organisatorische, onderwijskundige, psychometrische, en niet te vergeten acceptabiliteitseisen die aan herkansingsprocedures gesteld moeten worden, wordt niets gezegd. Dat is erg jammer, omdat juist deze herkansingsprocedure in een systeem van stringente ingangselectie, een zeer belangrijke rol kan gaan spelen, en dat zowel in een aantal gunstige aspecten, maar waarschijnlijk ook in andere opzichten die minder aanvaardbaar zijn. (een voorbeeld van het laatste kan zijn, dat de studieleiding zich van een stuk verantwoordelijkheid kan proberen schoon te wassen door te verwijzen naar de herkansingsmogelijkheden 'die dan toch voor iedereen die zich tekort gedaan voelt, openstaan'.)

Het grootste deel van zijn betoog hier, wijdt De Groot aan twee door hem onderscheiden modellen van rechtvaardiging, de predictieve, en de inhoudelijke rechtvaardiging. In zijn bespreking haalt hij een aantal zeer fundamentele psychologische principes overhoop, wat het onvermijdelijk maakt op dit deel van het rapport wat uitvoeriger in te gaan dan tot hier toe gebruikelijk was.

ad 1. de predictieve rechtvaardiging.

Tegen dit model heeft de Groot bezwaren.

"Ten eerste is dit model alleen overtuigend als het zuiver kan worden toegepast; en dat kan niet in een situatie van "maximale rechten". Er moet, om zo te zeggen, in deze doelmatigheidswijn veel billijkheidswater worden gedaan."

De Groot redeneert allerreerst vanuit de misvatting dat het predictiemodel empirisch ontwikkeld wordt: "Idealiter neemt men in de formule alle gegevens op, die de succespredictie kunnen versterken ongeacht hun inhoud met gewichten, die op empirische basis zó bepaald zijn, dat die versterking maximaal is." Hoewel niet ontkend kan worden dat deze karikaturale werkwijze van de psychometricus/ testconstructeur inderdaad wel eens in de praktijk wordt gevonden (en dat is te betreuren) is het toch zeker niet zo, dat dit de kenmerkende werkwijze is.

Ook wanneer we uitgaan van een reëler model dan De Groot schetst is het in het geheel niet duidelijk waarom het model alleen overtuigend zou zijn wanneer het 'zuiver' kan worden toegepast. In de literatuur zijn de gevallen overbekend waarin empiristische methoden tot grove discriminatie hebben geleid, reden waarom de meeste psychologen a priori deze benadering zullen afwijzen.

De werkwijze die dan wèl gevolgd wordt ligt in het zorgvuldig kiezen van de data die in het predictieve model gehanteerd zullen worden (dus geen shotgunapproach), niet veel anders dan het 'billijkheidswater' van de Groot, In concreto: predictieve validiteit op zich is niet voldoende, de procedure moet ook voldoende constructvaliditeit bezitten. Deze constructvaliditeit wordt weer gerelateerd aan de bedoelingen die men heeft met het selectieprogramma. Deze bedoelingen kunnen al dan niet in termen van acceptabiliteit geformuleerd zijn, dat maakt voor de toepassing van dit psychometrische model niet uit. De Groot lijkt dus vervallen te zijn in de fout het psychometrisch model zonder meer gelijk te stellen met de opvatting dat koste wat kost de predictieve validiteit zo hoog mogelijk moet worden opgevoerd; natuurlijk is dat een karikatuur en getuigt tenminste van een stuk ongeïnformeerde vooringenomenheid van De Groot. Dat de constructvaliditeit een in de praktijk moeilijker hanteerbaar begrip is dan predictieve validiteit betekent niet dat de psychometricus zich zonder meer aan de blinde actuariële methode zal overleveren.

"Ten tweeden veronderstelt deze manier van denken dat men kan bouwen op de resultaten van validiteitsonderzoek."

Dat validiteits onderzoek is inderdaad de grote moeilijkheid. Niet alleen is het zeer moeilijk uitvoerbaar, maar bovendien zal het een zeer arbeidsintensieve en kostbare zaak worden. waarvan de resultaten pas na vele jaren bekend (en dan waarschijnlijk al weer verouderd) zullen zijn. Dat betekent dat het predictieve model in zijn meer ideale vorm

moeilijk realiseerbaar zal zijn. In hoeverre dit model toch werkbaar gemaakt kan worden, ondanks een aantal onzekerheden die binnen de grenzen moeten blijven die acceptabiliteit en billijkheid er aan stellen, is een vraag waarop een antwoord gezocht zou moeten worden. Voor bepaalde situaties (een zeer groot deel van de kandidaten moet verondersteld worden 'geschikt' te zijn, en de capaciteit van de opleiding stelt geen grenzen aan aantallen opneembare studenten) kan een apriori argumentatie (gegrond op bijv. besliskundige principes) opgezet worden dat een aanpak via het predictieve model niet mogelijk is (en dat zal niet zo erg verbazen, de in het voorbeeld geschetste omstandigheden geven ook geen enkele aanleiding om een toegangsselectie uit te voeren)

Overigens betekent praktische onuitvoerbaarheid nog niet dat in principe het model niet toepasbaar is. Bijvoorbeeld is het mogelijk op grond van dit model de onjuistheid van bepaalde selectie methoden aan te tonen (tenminste plausibel te maken), wat op zich al zeer verdienstelijk zou kunnen zijn.

"Ten derde staan in het bijzonder differentiële beslissingen in de grenszone uiterst zwak".

Dit soort beslissingen zal altijd zwak blijven staan, welk model ook gehanteerd wordt. Volgens De Groot is het met het predictieve model in dit opzicht erg slecht gesteld omdat testprocedure en beslissingsformule voor degenen over wie beslist wordt, ondoorzichtig zijn. Op deze misvatting van De Groot is onder zijn eerste bezwaar al uitvoerig ingegaan. Daar kan nog dit aan toegevoegd worden: een wegingsformule (voor de diverse data die in het predictiemodel gehanteerd worden om tot selectiebeslissingen te komen) kan er inderdaad nogal ingewikkeld uitzien. Of de zin van deze formule in korte tijd aan een kandidaat uitgelegd kan worden is niet van zo groot belang, het gaat erom of aangetoond is dat de selectieprocedure werkt zoals ze verwacht wordt te werken, in dat geval kan in principe aan iedereen uitgelegd worden waarom een bepaalde wegings formule zo in elkaar zit, en niet anders. Niet name kan dat gaan om de vraag of hetgeen door een bepaalde test of toets gemeten wordt, inderdaad op enigerlei wijze voor de verdere studie van belang is. Ook van belang zijn gegevens als: bij een bepaalde testscore is er ongeveer die en die kans om met een bepaalde inzet het onderwijs programma goed af te sluiten; waarbij overwegingen van acceptabiliteit misschien voorschrijven dat de uiteindelijke beslissing op grond van dergelijke gegevens door de betreffende kandidaat zelf gemaakt moet worden.

De Groot stelt het predictieve model tegenover het inhoudelijke model. Zoals gezegd is dit gedeeltelijk een schijntegenstelling omdat uiteraard iedere selectieprocedure aan behoorlijke constructvaliditeit moet bezitten (aantoonbaar moet hebben). Daarnaast is het zeker niet zo, dat een bepaald model niet zou moeten voldoen aan eisen van predictieve validiteit. Wanneer bijv. aantoonbaar is van een bepaalde beslissingsprocedure dat de afgewezenen niet veel minder kans zouden hebben om de studie met redelijk succes af te sluiten dan de wél geaccepteerden, dan is dat model niet hanteerbaar. Het is bijv. niet ondenkbaar dat een dergelijk model tot stand gekomen is door acceptabiliteits normen te hanteren, waarbij het toch nog mogelijk is dat door het ontbreken van een redelijke predictieve validiteit de procedure als geheel als onbillijk (en daarmee onacceptabel) afgewezen moet worden.

ad 2. de inhoudelijke rechtvaardiging.

Het verhaal dat De Groot hier afsteekt is gewoon een omschrijving van eisen van construct validiteit (en/of inhoudsvaliditeit). Dat voor het onderzoeken van de constructvaliditeit (en bij het ontwerpen van de gehele procedure) vooral argumentaties van onderwijskundige aard gehanteerd zullen en moeten worden, zal niemand verbazen. Het is gevaarlijk om de nadruk te veel te leggen op 'inhouds validiteit'. zoals De Groot doet:

"De inhoudelijke rechtvaardiging komt erop neer, dat men hun relevantie aantoonst voor het studieprogramma waarvoor geselecteerd wordt, door een verwijzing naar hun inhoud. Dit is met name mogelijk bij scores op studietoetsen en andere gegevens betreffende leer en studieprestaties."

Maar al te gemakkelijk ontaardt dit in face validity, Het lijkt mij veel zinvoller van meet af aan de nadruk te leggen op constructvaliditeit, waarmee het accent verlegd wordt van subjectieve inhoudsvalidering naar de eis van positieve demonstratie van verbanden tussen toets (selectie procedure) en volgend onderwijs, (eventueel ook voorafgaand onderwijs). De simpele verwijzing naar de inhoud, wat De Groot ziet als middel om de relevantie aan te tonen (zie citaat), moet vervangen worden door onderzoek van de construct validiteit van zowel afzonderlijke tests en toetsen, als van de selectieprocedure als geheel.

Wanneer voldoende aanwijzingen voor een goede construct validiteit gevonden worden, is het inderdaad te verwachten dat ook de predictieve validiteit hoger zal zijn. (let wel, hoger, en niet hoog, zoals De Groot in een voetnoot (blz 72) stelt van studietoetsen die 'zorgvuldig op hun relevantie en onderwijskundige representativiteit' uitgezocht zijn. Ook wanneer inderdaad een hoge predictieve validiteit gevonden wordt, moet de vraag gesteld worden wat dat betekent, hoe dat geïnterpreteerd moet worden. Voorbeeld van interpretatie moeilijkheden levert dit citaat uit R.B. Cattell & H.J. Butcher: The prediction of achievement and creativity, 2728):

".. we should expect a traditional intelligence test administered to children in a particular year to correlate better with school achievement measured in the following year than would a culturefair test measuring gf, for the criterion that we are out to predict is already partly being measured in the test that sets out to predict it. But this slightly higher value in the correlation within a few months should not blind us to three considerations: (1) in predictions five to ten years hence, the culturefair test is likely to do much better; (2) the traditional test may be systematically unfair to individuals who do not happen to have had training in the particular skills in the preceding years relative to those who have such advantage; and (3) we obtain less psychological insight into what is happening in individual cases. Indeed, we are running in circles, by merely predicting what we already know!"

Dat betekent natuurlijk niet dat je van een dergelijke predictieve validiteit dan maar geen gebruik zou moeten maken, maar wel bijv. zijn dergelijke inzichten in hoe een bepaalde predictieve validiteit tot stand komt, van belang voor overwegingen t.a.v. de acceptabiliteit van een selectieprocedure.

Deze illustratie is ook hierom van belang, omdat De Groot niet alleen hier, maar ook op andere plaatsen in het rapport, veel belang hecht aan studie toetsen, die mits goed geconstrueerd ook een hoge predictieve validiteit zouden hebben. Daargelaten of die predictieve validiteit 'hoog' zal zijn, is het zeker zo dat de geciteerde kanttekeningen van Cattell en Butcher typisch op deze situatie van toepassing zijn. De Groot is er echter niet aan toe gekomen op genoemde vragen in te gaan, wat een groot gemis genoemd kan worden.

De Groot noemt in zijn toelichting als een belangrijk voordeel van het model van inhoudelijke rechtvaardiging, dat het daarbij voor de student

mogelijk is een strategie te volgen die hem met redelijke zekerheid zal doen slagen. Het is natuurlijk altijd aan te bevelen een selectieprocedure in deze zin op te zetten. Het predictieve model maakt dit dan ook zeker niet onmogelijk. Het vervelende is dat bij quota numerus fixus regelingen het hele verhaal niet opgaat, omdat dan de normen niet absoluut (criteriumgeoriënteerd) vastgelegd kunnen worden, maar noodzakelijkerwijs relatief zijn. Daarnaast, en misschien wel belangrijker, zal het altijd zo zijn dat bepaalde mensen, ondanks grote inspanningen, niet in staat zijn een van te voren vastgelegde absolute norm te halen. Je kunt dan zeggen dat deze kandidaten ook niet in het onderwijs thuis horen, maar daar gaat het niet om. Het punt is dit: op een intelligente test kun je je nauwelijks voorbereiden, en daarom mag volgens De Groot een dergelijke test niet in de testbatterij (het examen) voorkomen. Maar het verschil met een objectieve studietoets, met bekende slaagnorm, waar sommige mensen wel en andere niet aan kunnen voldoen, met dezelfde voorbereidende inspanning, dat verschil wordt dan wel erg klein. Uit een en ander volgt dat de redenering van De Groot niet zonder meer overgenomen kan worden (misschien wel gedeeltelijk opgaat, voor welk deel, dat zal in verdere discussie uitgemaakt moeten worden.)

Een acceptabel selectiecontract: transparantie

voetnoot op blz 82:

"In principe kan hier ook staan 'predictieve validiteit'. Op grond van de redenering in 5.4 lijkt het echter zeer onwaarschijnlijk, dat hiervoor een redelijk minimum te halen is."

De Groot is kennelijk bereid te werken met een selectieprocedure die een te lage predictieve validiteit heeft. Vergelijk wat hierover in dit stuk op blz 18 onderaan gezegd is. En vergelijk vooral met een andere voetnoot, blz 78:

"Men kan daar overigens als extra argument aan toevoegen, dat de algemene psychometrische ervaring leert, dat aldus zorgvuldig op hun relevantie en onderwijskundige representativiteit uitgezochte studietoetsen gewoonlijk óók een hoge predictieve validiteit bezitten."

Als het een waar is, kan het ander dat niet zijn, en omgekeerd. Over het algemeen opteert De Groot voor zijn blz 78 uitspraak. Hij stelt zich met een en ander wel bloot aan de beschuldiging dat hij lukraak wat

argumenten bij elkaar harkt als ondersteuning van zijn betoog en voorstellen.

"Een examen of andere selectieprocedure is, bij definitie, geheel transparant, als de gegadigde (examinandus) over alle informatie beschikt, die hij nodig heeft om de voor hem persoonlijk best mogelijke strategieën voor voorbereiding en deelname aan het examen te ontwikkelen." "Bij een strikte interpretatie van acceptabiliteit, in een situatie van maximale rechten, is een dergelijke volledige transparantie onmisbaar"(82).

Op blz. 20 zijn we hier al op ingegaan, daar bleek dat in bepaalde situaties (numerus fixus) de transparantie niet gehandhaafd kan worden, en dat ook andere moeilijkheden bij de interpretatie zich voordoen. Het is ongetwijfeld nodig bij iedere selectieprocedure aandacht aan de transparantie van een en ander te schenken, in de extreme vorm zoals De Groot die voorstelt zal zij echter vaak niet hanteerbaar blijven.

Op de toelichting van de Groot is al eerder geleverde kritiek van toepassing. Bijv., dat sommige niet objectieve procedures niet verenigbaar zijn met transparantie is evident, maar dat alle niet objectieve procedures in strijd zijn met de eis van transparantie is zeker niet evident.

In de laatste zin van deze paragraaf suggereert de Groot dat de problemen zich afspelen binnen één enkele test:

"het eenvoudigste systeem ieder item een punt is het meest transparante, dus het beste."

Bij deze selectie problematiek gaat het er niet om of een bepaalde .toets transparant is, maar moet de gehele selectieprocedure transparant zijn. En verondersteld mag worden dat transparantie van de selectieprocedure niet zonder meer te bereiken is door er voor te zorgen dat iedere van de onderdelen transparant is. Dit laatste wordt direct inzichtelijk wanneer men er van uit gaat dat volledige transparantie niet bereikbaar is, dat met name iedere afzonderlijke toets slechts onvolledig transparant kan zijn, en waarschijnlijk zelfs voor afzonderlijke studenten een verschillende transparantie zal hebben. Dat betekent dat de transparantie van een selectieprocedure een grootheid is waar alleen gissingen naar gemaakt kunnen worden.

Waar het gaat om selectie procedures met meerdere tests (toetsen) moeten waarschijnlijk andere maatregelen genomen worden om een zo goed mogelijke transparantie te bewerkstelligen. De problemen zijn groter dan in het geval van de enkele toets. Problemen van normen en waarden zullen zich voordoen, Bovendien zijn hier twee verschillende belangen in het spel, die ieder voor zich andere eisen aan de selectie procedure stellen : het belang van de instelling die de selectie uitvoert, en het belang van de individuele student die ook op de een of andere wijze een optimale strategie in zijn leven wil kunnen volgen. Mijns inziens heeft de eis van transparantie erg veel te maken met het laatste, de 'selectie' die het individu op zich zelf pleegt, en waartoe hij van zo goed mogelijke informatie voorzien moet worden; dit impliceert een ander psychometrisch model dan voor gewone institutionele selectie gebruikt wordt, en de vraag is in hoeverre het mogelijk is selectieprocedures te ontwikkelen waarin op de een of andere wijze beide belangen, van de individuele student zowel als van de selecterende instelling, zo goed mogelijk aan bod komen.

5.6 Schets van een acceptabel selectieexamen.

Een aantal selectiemiddelen sluit de Groot uit, met als gevolg dat alleen objectief scoorbaar examenwerk nog acceptabel kan zijn, mits het voldoet aan een aantal eisen, o.a. t.a.v. transparantie, objectiviteit, psychometrische kwaliteiten (blz. 84)

Zowel m.b.t. institutionele selectie als m.b.t. optimalisering van individuele keuzestrategieën van de student, is het mijns inziens gewenst om gebruik te maken van beschikbare of in principe beschikbaar te maken informatie, wanneer deze informatie een duidelijke relevantie heeft voor de selectie beslissing. Zo zou bij vooronderzoek kunnen blijken dat de volgende gegevens van belang zijn: biografische gegevens, persoonlijkheids en intelligentie gegevens, motivatiegegevens, schoolcijfers, referenties. Dat dergelijke gegevens in bepaalde opzichten strijdig kunnen zijn met acceptabiliteits eisen, wil nog niet zeggen dat ze zonder meer uitgesloten moeten worden. Bijvoorbeeld worden dergelijke gegevens gebruikt bij counseling van individuele studenten; m.a.w. het zou zinnig zijn om hier enige systematiek te brengen, te kwantificeren, in wat i.h.a. vaag, kwalitatief, blijft; en om valide interpretaties op basis van dergelijk soort informatie aan iedere student ter beschikking te stellen, en niet alleen aan diegenen die zich bij de studenten dekaan of psycholoog of psychiater aanmelden. Een duidelijke relevantie heeft deze informatie

dus zeker t.a.v. de door de individuele student te nemen beslissingen. Maar ook t.a.v. de institutionele selectie is het belang evident, en wanneer men er voor kiest, op welke principiële of pragmatische gronden dan ook, om van dergelijke informatie geen gebruik te maken, moeten toch de kosten en nadelen die uit een dergelijke handelwijze voortvloeien, afgewogen worden tegen de voordelen die verondersteld worden behaald te worden.

Met andere woorden: sommige selectiemiddelen blijken niet geheel acceptabel te zijn, het zou echter heel goed mogelijk kunnen zijn dat het niet gebruik maken van deze middelen ook leidt tot consequenties die in bepaalde opzichten niet acceptabel zijn. Wanneer men aan een bepaalde procedure een foutje ontdekt, kan men niet zonder meer besluiten tot het niet volgen van deze procedure, maar de voor en nadelen van de ene werkwijze moeten tegen die van andere. alternatieve, selectieprocedures afgewogen worden.

5,7 validiteit van studietoetsen.

De Groot beantwoordt hier allereerst de vraag of het, gegeven een studieprogramma waarin het in hoofdzaak gaat om "cognitieve (c.q. theoretische) vaardigheden". Mogelijk is een examen te construeren dat, gegeven dat voldaan is aan eisen van billijkheid en constructvaliditeit, ook voldoet aan eisen van predictieve validiteit.

"de predictieve validiteit van 'zorgvuldig op hun relevantie en onderwijskundige representativiteit uitgezochte studietoetsen' en vergelijkbare andere examenonderdelen (5.6) blijkt gewoonlijk hoog te zijn, gemiddeld hoger dan van welke ander type voorspellingsmiddel ook."

Dit is een triviale opmerking. Immers, naar de mate waarin een test meer het karakter krijgt van een steekproef uit het te voorspellen gedrag, in die mate zal ook de predictieve validiteit(t.a.v. dat gedrag !) hoger zijn.

De Groot neemt de term 'predictieve validiteit' in een niet ongebruikelijke, maar toch zeker beperkte betekenis, nl. voorspelling van studiesucces. Zijn inleidende babbel in de eerste hoofdstukken van het rapport zou doen vermoeden dat hij ook zijn gedachten zou laten gaan over de predictieve validiteit t.a.v. de beroepsprestaties ná het voltooien van de studie. De vraag dringt zich namelijk op of studietoetsen, die misschien

heel erg valide kunnen zijn binnen de onderwijssituatie, ook nog valide kunnen zijn in voorspelling van beroepssucces. De gegevens die bekend zijn over verband tussen studieprestaties en beroepsprestaties zijn niet al te hoopgevend (zie bijv. Gronbach Essentials etc. 1970), en inderdaad kunnen er op onderwijskundige en andere gronden een groot aantal argumenten gegeven worden waarom een dergelijke predictieve validiteit laag zou zijn of misschien zelfs niet aantoonbaar.

De Groot laat zich niet uit over hoe hoog die predictieve validiteit (t.a.v. studiesucces) zou kunnen zijn. Het is waarschijnlijk zo dat, ondanks de waarschijnlijkheid dat een studietoets een hogere pred. val. heeft dan ieder ander test, deze validiteit geen hogere waarde dan .50 of 55 bereikt (en dat zou dan nog een heel erg ideaal geval zijn) De Groot onttrekt zich aan de opgave om te berekenen (in de zin zoals Hazewinkel dat bijv. gedaan heeft) of selectie in de situatie waar De Groot hier vanuit gaat, zelfs in het geval van een relatief hoge pred. val., zinvol is d.w.z. significant (in statistische zin of financiële zin) beter dan geen selectie of random selectie)

"Goed geconstrueerde studietoetsen zijn met name vrijwel altijd betere voorspellers van studieprestaties (of intellectueel studiesucces) op het cognitieve gebied dat zij bestrijken, dan intelligentietests en speciale voor predictiedoeleinden geconstrueerde geschiktheidstests." (86)

Hier is weer relevant het op blz 19-20 gegeven citaat uit Cattell en Butcher 1968,

Volgens De Groot zijn studietoetsen des te betere voorspellers naarmate "de te testen kandidaten verder gevorderd zijn (hogere leeftijd, hoger niveau) zodat de vereiste cognitieve vaardigheidsfactoren zich al voor een deel hebben kunnen uitkristalliseren (d.i. beter meetbaar en voorspelbaar hebben kunnen worden)"

Uit de psychologie is mij bijzonder weinig bekend over "uitgekristalliseerde" cognitieve vaardigheidsfactoren. 'Waarschijnlijk bedoelt De Groot te zeggen dat degenen die zich aanmelden voor hoger onderwijs al een zeer lang en intensief selectieproces achter de rug hebben, waar vooral geselecteerd is op deze cognitieve vaardigheidsfactoren. Dat betekent echter niet dat studietoetsen die kapitaliseren op deze cognitieve vaardigheidsfactoren (en dan in feite fungeren als intelligentie tests !) studiesucces dan beter zouden voorspellen. Integendeel! De voorspelbaarheid op grond van cognitieve

vaardigheidsfactoren van studiesucces in het hoger onderwijs zal juist zeer slecht zijn omdat er een zeer sterke restriction of range op deze variabelen bestaat (anders dan bijvoorbeeld in Amerika aan het eind van de highschool periode het geval is) Een voorbeeld van wat het over het hoofd zien van dergelijke restriction of range effecten tot gevolg kan hebben is te vinden in de dissertatie van Willems, 1959, en zijn artikel "Een proefschrift beproefd", Neds T. vod. Ps., 1964.124135,(sorry, ik heb het mis, in dit artikel wordt niet gesproken over restriction of range, hoewel dit wel eens verband zou kunnen houden met de onverwachte resultaten die 'Willems verkreeg)

In plaats van te verwijzen naar de "zee van onderzoeksgegevens'¹ had de Groot er beter aan gedaan man en paard te noemen, zoals bijv het artikel van Cronbach over validiteit in R.L. Thorndike (Ed:) Educational measurement, 1971, en het overzicht van Cattell en Butcher The prediction of achievement and creativity, 1968.

De Groot twijfelt er niet aan dat een dergelijk examen als in deze paragraaf bedoeld, een zeer valide selectiemiddel kan zijn. (87)

De gegeven commentaar betekent, om het even kort te recapituleren, dat deze conclusie van de Groot voorbarig en waarschijnlijk niet juist is; gronden zijn: 1. de beperkte zin waarin de term predictie gehanteerd wordt (nl. slechts van studiesucces) 2. de triviale situatie, waarin examen een proef is van zoveel mogelijk dezelfde inhoud als het er op te volgen onderwijs (vgl. selectie van typistes door een bedrijf, waar een eenvoudige vergelijkende proeve van bekwaamheid mogelijk is) 3. het 'in omgekeerde zin' gebruiken van het restriction of range effect, als iets dat voorspelling beter mogelijk maakt i.p.v. slechter; 4, het verzuim om kort de effectiviteit van deze selectie te berekenen, uitgaande van een aantal redelijkerwijs te maken veronderstellingen, zoals bijv. eerder door Hazewinkel gedaan.

de groot 26

"Pleidooien voor cognitieve tests naast een goed selectieexamen staan praktisch altijd, en zeker op het niveau van het hoger onderwijs, ook uit een oogpunt van doelmatigheid (predictieve validiteit) uiterst zwak. Voor dit doel zijn ze niet nodig."(88)

De Groot geeft voor deze uitspraak een volkomen willekeurig, uit de lucht gegrepen argumentatie (blz 87-88), en verwijst ook niet naar relevante

literatuur waarop zijn argumenten gebaseerd zouden kunnen zijn.

Cattell and Butcher (1998): "... tests of abilities alone are probably not the best instrument for the prediction of scholastic success at this level (universitair onderwijs). This conclusion is suggested by a large number of studies, and particularly by a careful and thorough survey made by Sanders (1961), who compares British, American, and Australian methods of university selection. Even at the university selection level, abilities have not become anything other than relatively less important, and hardly even that."

(dit dus in vergelijking tot de situatie voor middelbaar onderwijs:
"Experience from studies of selection for secondary schools in Britain strongly suggests that a test of general intelligence, given at age eleven, is more effective in predicting school attainment than tests in English or arithmetic.")

"..... what little evidence is available suggests that tests of ability by themselves (onderstrepen van mij) are not very effective in predicting student grades in British universities, partly, no doubt, because of the highly selected nature of the population."

Als je er vanuit gaat dat er een goed selectieexamen al is, zoals de Groot doet (zie boven), dan heeft toevoeging van intelligentie tests hoogstwaarschijnlijk weinig zin t.a.v. het verbeteren van de validiteit van de selectie. Hierbij moeten twee opmerkingen gemaakt worden. Ten eerste is het voorlopig hoogst twijfelachtig dat voor het universitaire onderwijs goede selectie procedures ontwikkeld kunnen worden (bijv. op basis van alleen studietoetsen); in deze gevallen zal het toevoegen van tests op cognitief gebied een verbetering van de kwaliteit op kunnen leveren. Voor het speciale geval van selectie voor de universiteit bestaat er, bij de huidige vorm van het middelbaar onderwijs in ons land, echter geen reden om hier enig belangrijk effect te verwachten aangezien het studenten aanbod juist op eigenschappen als intelligentie al in belangrijke mate in het voorafgaande onderwijs geselecteerd is. (restriction of range beperkt het voorspellende gebruik dat van de betreffende data gemaakt kan worden). In zoverre is de conclusie van de Groot wèl juist. Ten tweede leidt het kunnen beschikken over intelligentietest resultaten er toe dat de selectie procedure inzichtelijker wordt op het punt van zijn constructvalidering. Ook wanneer gegevens over intelligentie op zich de voorspelling van bijv. studiesucces niet kunnen verbeteren, vormen zij toch een controle middel op de werking

van de selectieprocedure. Vanuit dit oogpunt is het altijd raadzaam om intelligentie gegevens e.d. te verzamelen wanneer daartoe de mogelijkheden bestaan.

met betrekking tot persoonlijkheids variabelen.

"De desbetreffende tests en vragenlijsten mogen voor doeleinden van studie counseling een zeker belang hebben, hun validiteit en gerapporteerde correlaties met, en het daaruit af te leiden systematische belang voor latere studieprestaties zijn voor selectie doeleinden oninteressant."(88)

Alle gegevens die van 'systematisch belang voor latere studieprestaties' zijn, zijn ipso facto van belang voor de selectieve beslissings strategie die men kiest.

Ofwel er is een systematisch verband met later studiesucces, en in dat geval moeten persoonlijkheids variabelen een rol spelen bij de selectie (gewild of ongewild);

ofwel er is geen systematisch verband met later studiesucces, en dan zijn deze variabelen ook voor selectie niet van belang. Allebei tegelijk, zoals de Groot het kennelijk ziet, is niet mogelijk, In de voetnoot op blz 89 geeft de Groot weer een staaltje van averechtse logica weg, daarmee opnieuw illustrerend dat de door hem gegeven argumentatie van nul en generlei waarde is:

"Men kan daaraan toevoegen, dat wat aan deze factoren relevant is, evenals dat bij de intelligentie het geval was, in een studietoets over bestudeerde stof voor een belangrijk deel impliciet wordt meegemeten: en wel. als het goed is, betrouwbaarder, specifiekere en relevanter dan met enige andere test mogelijk zou zijn." (onderstreping van mij).

Dat je iets, wat je alleen maar impliciet meet, ook nog betrouwbaar etc. zou meten, is baarljke nonsens. Wanneer onbekend is in welke mate je iets meet, en hoe je dat meet, (en dat is de situatie bij 'impliciet meten'), ontbreekt uiteraard ieder inzicht in meeteigenschappen als betrouwbaarheid, validiteiten etc.

Om nog maar niet te spreken van de opvallende constructie dat een dergelijke impliciete meting ("als het goed is") een betere meting zo,) zijn dan een meting op grond van een expliciet voor dit doel ontwikkelde test!

In zijn toelichting op deze paragraaf houdt de Groot een klets verhaal over intelligentie dat zo warhoofdig in elkaar zit, onzinnige uitspraken naast de meest triviale gemeenplaatsen bevat, dat ik mij niet genoodzaakt voel daar nog verder op in te gaan. Twee opmerkingen toch nog: (1) waar hij op blz 83 nog stelde dat Intelligentie tests onaanvaardbare selectie middelen zijn, is hier zijn verhaal één groot pleidooi voor selectie op grond van cognitieve factoren. Hij kan dat doen op grond van de zojuist gekritiseerde onwaarschijnlijke gedachtengang dat sommige variabelen door niets zo goed gemeten worden als door tests die niet ontwikkeld zijn om ze te meten; in dit geval:

"Als wij selecteren op cognitieve factoren, dan moeten wij dat dus in de allereerste plaats doen op die algemene vakfactoren (g), op die 'gespecialiseerde intelligentie'. Dat is precies wat men in een goed opgesteld selectieexamen doet; men doet dat in het bijzonder als de selectie plaats vindt na een jaar (algemene en) vakpropeddeuse."(90)

(2) "Toekomstige studiestudieprestaties zijn op grond van niets zo goed te voorspellen als die van vroegere studiestudieprestaties."

Het is inderdaad waarschijnlijk dat dit ook voor selectie voor het hoger onderwijs zal gelden. Op zich betekent dat echter nog niet dat je dan juichend moet concluderen dat dus selectie d.m.v. studietoetsen het enige en beste middel is om te selecteren voor een universitair onderwijs dat zich ook verder kenmerkt door veelvuldige studietoetsen. Integendeel, dit is volkomen blindelinge selectie, actuariële predictie in zijn slechtste vorm.

5,8 Selectie op nietcognitieve (praktische) vaardigheden.

Het is niet duidelijk waarom de Groot een zo uitgebreide aparte paragraaf hieraan besteedt. Voor het opnemen van variabelen als deze in de selectieprocedure moet precies dezelfde weg gevolgd worden als voor andere, bijv. cognitieve, variabelen. Bij beslissingen of bepaalde variabelen bij de selectie een rol moeten (mogen) spelen mag men in ieder geval niet afgaan op vage noties en persoonlijke opvattingen, zoals de Groot in deze paragraaf ten beste geeft. Afzonderlijke commentaar behoeven de voorwaarden die de Groot noemt, wil het meetellen van praktische vaardigheden bij selectie acceptabel zijn. (en deze voorwaarden zouden dan ook voor andere variabelen moeten gelden, al

zegt de Groot dit niet expliciet).

Meetellen is acceptabel als deze variabele:

"(1) niet door iedere overigens aannemelijke student, met voldoende oefening tot een voldoende niveau te ontwikkelen is.

(2) niet zonder bezwaar voor opleiding en beroep te compenseren of te omzeilen is;

(3) als aparte aanlegfactor voldoende betrouwbaar kan worden gemeten;

(4) aldus gemeten een goede voorspeller oplevert voor het aanwijzen van degenen die het "nooit zullen leren" (validiteit);

(5) ook na de te verwachten zelfselectie bij de aanmelding voor deze studie, bij een substantieel percentage van de te selecteren studenten als onvoldoende moet worden gekwalificeerd." (96) Bovendien eist de Groot dat al deze condities 'in voldoende mate vervuld' zijn.

ad 1. ook als dit wel het geval is, moet de vraag gesteld worden of het mogelijk is (de bereidheid bestaat, het de bedoeling is) dat in het onderwijs de bedoelde vaardigheden geleerd worden.

ad 2. allicht, in het andere geval is deze variabele niet van belang voor een ingangselectie.

ad 3. als iets niet meetbaar is (of met onvoldoende betrouwbaarheid en (construct)validiteit gemeten wordt) kan het uiteraard niet meespelen in een selectie procedure

ad 4. Het gaat hier niet om constructvaliditeit (zoals in 3) maar om predictieve validiteit (wat overigens iets anders is dan de mate waarin "degenen die het 'nooit zullen leren'" aangewezen kunnen worden). Als de bijdrage in predictieve validiteit die een bepaalde variabele levert xxx bij toevoeging aan de selectiebatterij onvoldoende is, ja, dan nemen we hem uiteraard niet in de selectie procedure op. Punt 2,3, en 4 zijn dus trivialiteiten.

ad 5. Op dit punt geeft de Groot nog een afzonderlijke toelichting van bijna 2 bladzijden. Deze komt neer op de, op zich volkomen juiste, demonstratie dat in situaties waarin aangenomen moet worden dat het percentage 'ongeschikten' zeer klein is, pogingen om deze ongeschikten uit te selecteren resulteren in een nogal fors percentage foute beslissingen onder de afwijzingen. De fout die de Groot maakt, is dat hij deze redenering toepast op afzonderlijke onderdelen van een selectie batterij (in dit geval praktische vaardigheid), terwijl deze redenering alleen op de totale selectie procedure mag worden toegepast.