

SELECTIEPROBLEMEN

BIBLIOTHEEK-NEDERLANDS INSTITUUT
VOOR PRAEVENTIEVE GENEESKUNDE
WASSENARSEWEG 56 - LEIDEN

DRUK: N.V. DRUKKERIJ V/H BATTELJEE & TERPSTRA — LEIDEN

UJCF
W 62
1)

PUBLICATIE VAN HET NEDERLANDS
INSTITUUT VOOR PRAEVENTIEVE GENEESKUNDE

SELECTIEPROBLEMEN

DE BEOORDELING VAN GESCHIKTHEID VOOR FUNCTIE,
STUDIE EN BEROEP

DR. S. WIEGERSMA



1964

SWETS & ZEITLINGER/AMSTERDAM

BIBLIOTHEEK-NEDERLANDS INSTITUUT
VOOR PRAEVENTIEVE GENEESKUNDE
WASSENAARSEWEG 56 - LEIDEN

Deze uitgave is een bewerking van de in „Mens en Onderneming” in 1962 (afl. 4 en 6)
en in 1963 (afl. 1 en 2) verschenen artikelen „Selectieproblemen”.

Copyright 1964

INHOUD

	pag.
I. GRONDBEGRIPPEN	8
Inleiding	8
Begripsbepaling	8
De situatie	8
De toetsing	9
De persoon	10
Het criterium	11
Het operationeel maken van een criterium	12
Validatie	15
II. PREDICTOREN	17
Inleiding	17
De keuze van de predictoren	17
Verificatie van een keuze	18
Eisen, aan de predictoren te stellen	19
De kansverwachting van de geschiktheid	22
III. EFFECTIVITEIT EN RENDEMENT	30
Inleiding	30
Geschiktheid voor v.h.m.o.	30
Cut-Off-scores	31
Correlatie	35
Correlatie en effectiviteit van selectie	36
Het rendement van de selectie	41
IV. OPERATIONS RESEARCH	45
Inleiding	45
Keuze tussen onderwijsmogelijkheden	45
Operations research	50
Sequente selectie	50
Distribuerende selectie	52
V. VALIDITEIT	54
Inleiding	54
De invloed van selectie op de normstelling	54
Validiteit	56
Practische evaluatie van de validiteit	56
LITERATUUR	59
SUMMARY	60

Selectieproblemen

Er is een tijd geweest, waarin toegepaste psychologie vrijwel identiek was met 'testen', terwijl dit testen zeer overwegend geschiedde in situaties, waarin geselecteerd moest worden. Psychologische personeelselectie, recrutenkeuringen, testonderzoek bij beroepskeuzevoorlichting, zijn hiervan bekende voorbeelden. In de tijd, waarin zij nog nieuwe verschijningen waren, zijn zij heftig omstreden geweest: men 'geloofde' in hun waarde, of men 'geloofde er niet in' en in beide gevallen had men weinig mogelijkheden om zijn overtuiging anders dan door voorbeelden te argumenteren.

Immiddels is de ontwikkeling verder gegaan. Het 'testen' vond ruimer toepassing dan alleen bij selectie en het arbeidsveld van de toegepaste psychologie werd geleidelijk veel ruimer dan 'alleen maar testen'. Een vergelijking van werken over bedrijfspsychologie is in dit opzicht illustratief: nog kort na de oorlog handelden zij voornamelijk over personeelselectie, terwijl moderne publicaties daarentegen vol staan met beschouwingen over personeelsverhoudingen, 'counseling', veiligheid en ergonomie, zelfs zodanig, dat aan selectievraagstukken soms nauwelijks aandacht wordt gewijd. De psychologische personeelselectie is intussen maatschappelijk gearriveerd, zij wordt in brede kring als vanzelfsprekend aanvaard en is daardoor ogenschijnlijk niet meer zo interessant als de modernere en meer problematische onderwerpen.

Toch bedriegt hier de schijn enigszins. Het selectievraagstuk — in werkelijkheid niet één, maar een gehele verzameling van vraagstukken — is nog allerminst opgelost. Nog steeds wordt ook aan deze problematiek gewerkt, waarbij één van de belangrijkste ontwikkelingen is geweest, dat men heeft ingezien, dat de psychologische selectie slechts een bijzonder geval is van een veel algemener selectievraagstuk. Het lijkt daarom de moeite waard, om de huidige stand van dit vraagstuk eens in een serie beschouwingen te belichten. In het eerste hoofdstuk, dat hierna volgt, zullen voornamelijk enkele grondbegrippen worden behandeld. Als meest essentiële punten komen daarbij de betekenis van de selectiecriteria en het belang van validering van selectieresultaten naar voren.

Het tweede hoofdstuk zal handelen over de predictoren, die wij gebruiken om een toekomstige ontwikkeling te voorspellen. Nadat een aantal eisen is besproken, waaraan predictoren hebben te voldoen, zal de relatie tussen predictor en criterium nader worden gezien. Dit geschiedt aan de hand van een aantal concrete voorbeelden.

In het derde hoofdstuk wordt eerst, opnieuw aan de hand van een voorbeeld, gedemonstreerd, op welke wijze op grond van een predictor beslissingen kunnen worden genomen. Daarna komt het probleem van de effectiviteit van de selectie aan de orde, met daaraan gekoppeld het vraagstuk van het rendement van selectie.

In het vierde hoofdstuk wordt aandacht besteed aan een moderne ontwikkeling, die wel betiteld wordt als 'operations research'. Het laatste hoofdstuk behandelt de consequenties van selectie, met name de vraag, in hoever selectie het beoordelingssysteem beïnvloedt en daardoor zichzelf ondergraaft. Aansluitend hierop komt dan tenslotte het probleem van de validiteit van selectie nogmaals aan de orde.

HOOFDSTUK I

GRONDBEGRIPPEN

Inleiding

Op tal van punten en plaatsen in het leven worden wij voor opgaven geplaatst, waarin een beslissing moet worden genomen over de plaatsing van mensen in situaties. Er dient dan een selectie plaats te vinden, om aan te wijzen, welke persoon of situatie het beste aan een bepaalde norm beantwoordt.

Vaak zal deze selectie intuïtief of misschien zelfs onbewust geschieden, maar er zijn ook gevallen, waarin men er althans naar streeft om rationeel en methodisch te werk te gaan. Op deze laatste heeft de hierna volgende reeks beschouwingen betrekking. Het is niet de bedoeling, daarbij in te gaan op de praktijk van de selectie, dus bijv. op de vraag, welke concrete methoden of hulpmiddelen kunnen worden gebruikt om in een specifiek geval een zo effectief mogelijke selectie uit te voeren. Onze doelstelling is uitsluitend, een aantal principiële aspecten van het selectievraagstuk te belichten. Ter illustratie zullen voorbeelden worden gebruikt, die ontleend zijn aan de psychologische praktijk; ook buiten de psychologie kan men dezelfde principes echter evenzeer toepassen.

Begripsbepaling

Onder *selectie* verstaan we: *de toetsing, met behulp van een geschiktheids-criterium, van de relatie tussen een gegeven persoon en een bepaalde situatie, met het oog op het maken van een keuze.*

Daar wij ons beperken tot rationele selectie, kan bij 'plaatsing in een bepaalde situatie' vooral worden gedacht aan de bezetting van bepaalde vacatures in een bedrijf, de toelating van leerlingen tot een school of andere opleiding, de differentiatie tussen verschillende beroepen, functies of opleidingen die voor een gegeven persoon open staan en dergelijke.

In de definitie komen vier elementen voor, die een essentiële rol hebben: toetsing, criterium, mens(en) en situatie. Aan ieder van deze elementen dient afzonderlijk aandacht te worden besteed. Buiten beschouwing blijft hier de subjectieve problematiek, die vrijwel steeds aan het maken van een keuze inherent is, terwijl ook niet zal worden ingegaan op de vraagstukken die bij de plaatsing zullen optreden.

De situatie

De concrete aard van de situatie waarvoor geselecteerd moet worden heeft uiteraard grote praktische betekenis in verband met de keuze van de selectiemiddelen. Zij is echter ook in verband met de gezichtspunten die aan de selectie ten grondslag worden gelegd van belang. Bij personeelselectie zal men bijv. dikwijls streven naar toelating van een zo gering mogelijk aantal

ongeschikten, terwijl bij schoolselecties veelal eerder geldt, een zo gering mogelijk aantal geschikten af te wijzen. Deze beide gezichtspunten leiden in het algemeen niet tot dezelfde beoordelingswijze, omdat met de toelating van een groter percentage geschikten gewoonlijk ook het percentage toegelaten ongeschikten toeneemt en met de afwijzing van een groter deel der ongeschikten tevens meer geschikten worden afgewezen.

Factoren die zwaar wegen voor een persoon die zich aanmeldt voor een vacature in een bepaald bedrijf, bijv. de vraag of elders wellicht mogelijkheden zijn die nog aantrekkelijker zijn, tellen voor het bedrijf niet, terwijl een bedrijf op zijn beurt weer rekening houdt met aanbiedingen van verschillende personen. Daardoor kan éénzelfde geval dus eventueel meer dan één selectiesituatie inhouden, waarvoor verschillende keuzeprincipes gelden. Op deze punten zullen wij later terugkomen. Daarnaast zullen er echter ook aspecten van het selectieprobleem blijken te zijn, waarvoor de aard van de situatie geen verschil maakt.

In verband met de situatie is het nog nuttig, onderscheid te maken tussen functie en plaats. Met de term *functie* duiden wij aan, een omlijnd geheel van taken en opgaven; dit begrip gebruiken wij dus ook, wanneer men normalerwijze van een beroep of van een opleidingsprogramma zou spreken. Het begrip '*plaats*' duidt een vacature of iets analoogs aan.

Per functie kunnen dus een verschillend aantal plaatsen beschikbaar zijn, waarover in het selectieproces geoordeeld wordt. Een aantal principiële aspecten van het selectieprobleem wordt pas goed kenbaar en zeker ook eerst praktisch van belang, wanneer het aantal plaatsen per functie betrekkelijk groot is.

De toetsing

Het uitgangspunt bij de selectie kan zijn een gegeven situatie of een gegeven persoon. Ten aanzien van vele principiële aspecten van het selectieprobleem maakt dit geen verschil: tal van inzichten blijven geldig, wanneer men de begrippen persoon en situatie systematisch verwisselt. Dit kan men het *dualiteitsprincipe van de selectie* noemen. Om te vermijden dat in het volgende telkens een vermoeiende dubbele formulering zou moeten worden gebruikt, wordt in het algemeen alleen van de beoordeling van personen ten opzichte van een situatie gesproken. De lezer bedenke echter dat in wezen steeds daarachter zou moeten volgen 'of van de beoordeling van situaties ten opzichte van een persoon'.

In het selectieproces moet nu beoordeeld worden, of en in hoeverre een gegeven persoon (c.q. een aantal personen) in verband met een bepaalde functie beantwoordt aan een norm, die aan het criterium wordt ontleend. Deze beoordeling kan op twee wijzen geschieden: absoluut en relatief.

Bij de *absolute beoordeling* beschikken we over één minimum-normwaarde, die differentieert tussen acceptabel en niet-acceptabel, tussen 'geschikt' en 'ongeschikt'.

Bij de *relatieve beoordeling* beschikken we over verschillende normwaarden, die in een onderlinge rangorde staan. Dit is vooral van belang, als we verschillende personen onderling willen vergelijken en rangschikken, bijvoorbeeld op hun geschiktheid voor een bepaalde functie. Denkbaar is ook een combinatie van de twee principes, bijv. doordat bij degenen die volgens de absolute maatstaf acceptabel zijn nog een relatieve beoordeling plaats vindt, of doordat na een onderlinge rangschikking een scheidingslijn tussen acceptabel en niet-acceptabel wordt getrokken.

De persoon

Op grond van de relatie tussen persoon en functie kunnen we vier typen gevallen onderscheiden:

- a. Eén persoon wordt beoordeeld voor één functie (*individuele selectie*). Dit geval doet zich niet alleen voor, wanneer één kandidaat wordt voorgedragen voor benoeming in een functie in het bedrijfsleven, maar ook bijv. wanneer een aantal leerlingen wordt beoordeeld voor toelating tot een school die voldoende plaatsen heeft om alle 'geschikten' op te nemen. Bij individuele selectie kan men volstaan met een absolute toetsing.
- b. Eén persoon wordt beoordeeld voor verschillende functies (*differentiatie*), bijv. in verband met school- of beroepskeuze of in verband met een sollicitatie of promotie in een bedrijf, dat verschillende mogelijkheden voor hem heeft. In dit geval moet primair een relatieve toetsing plaats vinden, eventueel gevolgd door een absolute beoordeling.
- c. Verschillende personen worden in verband met één functie beoordeeld, terwijl het aantal beschikbare plaatsen kleiner is dan het aantal kandidaten (*vergelijkende selectie*, '*Konkurrenzauslese*'). Ook in dit geval is er weer primair sprake van een relatieve, eventueel gevolgd door een absolute toetsing. Indien het aantal beschikbare plaatsen tenminste even groot is als het aantal personen, hebben we in feite met type a te maken.
- d. Verschillende personen worden ten opzichte van verschillende functies beoordeeld, terwijl per functie het aantal plaatsen veelal kleiner is dan het totale aantal personen (*distribuerende selectie*). Dit type selectie geeft het meest algemene geval weer, in principe omvat het ook de voorgaande typen. Evenals in type c vindt ook hier in eerste instantie een relatieve en absolute toetsing plaats, die per functie geschiedt.

Daarna dient echter nog volgens bepaalde richtlijnen een optimale distributie over de functies plaats te vinden. Daarbij kunnen zich tal van problemen voordoen. Het is bijv. niet uitgesloten dat iemand, die voor functie A de eerste plaats in de rangorde inneemt en voor functie B de derde plaats, toch bij voorkeur in deze laatste functie dient te worden geplaatst, omdat hij daar voor het geheel van de organisatie het meest waardevol wordt geacht. Het verwezenlijken van een optimale distributie is, wanneer het enigszins aanzienlijke aantallen functies en plaatsen betreft, een zeer

gecompliceerde opgave. Met name is het probleem, hoe de verschillende mogelijke distributies onderling vergeleken kunnen worden, nog lang niet opgelost. Voor grote organisaties is de oplossing van dit vraagstuk echter van zo eminent belang, dat een belangrijk deel van het theoretisch onderzoek over selectieproblemen hieraan gewijd is.

Het criterium

In het kader van selectie pleegt men dikwijls slechts weinig bewuste aandacht aan het selectiecriterium te besteden, blijkbaar omdat men impliciet veronderstelt, dat hierover geen verschillen van inzicht bestaan. Zolang de uitvoering van de selectieprocedure in handen is van dezelfde persoon, die ook de keuze moet doen en die bovendien achteraf de resultaten van de selectie moet gebruiken, is deze geringe mate van bewustheid nog wel aanvaardbaar. Dit wordt echter anders, wanneer de selectieprocedure, de keuzebeslissing en het gebruik van de selectieresultaten telkens door andere personen geschieden, bijv. de selectieprocedure door een arts of psycholoog, de keuze door een personeelchef, het gebruik door een bedrijfsleider. Dan kan het gemakkelijk gebeuren, dat ieder van hen een ander criterium hanteert, waardoor men tot tegenstrijdige conclusies komt.

Dit is te gemakkelijker mogelijk omdat er zeer gangbare criteria van sterk uiteenlopende aard bestaan: economische, psychologische, somatische, medische, sociale, etc. Men kan bijv. uitgaan van het verlangen naar een maximaal rendement van de geïnvesteerde middelen (*economisch*), of zo weinig mogelijk geschikten een kans willen onthouden (*sociaal*), of schade aan de gezondheid willen voorkomen (*medisch*), of ieder mens een optimale gelegenheid willen geven tot zelfontplooiing (*pedagogisch*), of degenen willen selecteren die een maximale prestatie kunnen leveren (*somatisch c.q. psychologisch*), of de samenwerking in een arbeidsgemeenschap zo goed mogelijk willen laten functioneren (*organisatorisch*).

Ook binnen één beoordelaar kunnen deze criteria wel tot botsingen leiden: de meeste mensen laten zich niet ten allen tijde uitsluitend door één en hetzelfde gezichtspunt leiden. Steeds bestaat dan de kans, dat verschillende criteria tot verschillende beslissingen zullen voeren.

In feite geldt, dat zodra er verschillende criteria in het geding zijn, een objectieve benadering van het selectievraagstuk niet meer mogelijk is. Alleen subjectief kan ieder mens voor zichzelf een bepaalde hiërarchie van normen vaststellen en met behulp daarvan trachten een oplossing te vinden voor een bepaald geval van strijdigheid tussen selectiecriteria.

Als bijv. iemand voor een bedrijf het meest rendabel is in een bepaalde zeer gespecialiseerde functie, die hem weinig mogelijkheden voor zelfontplooiing biedt, terwijl deze mogelijkheden wel aanwezig zijn in een andere functie, die echter voor het bedrijf economisch minder waardevol is, zal men voor zichzelf moeten uitmaken, in hoeverre er in economisch opzicht een offer zou kunnen worden gebracht om de betrokkene een betere ontplooiingskans te

geven. In een dergelijk geval dient men echter te voorkomen, dat de leiding van een bedrijf bijv. een economisch en de selecteur een pedagogisch-antropologisch criterium hanteert, zonder dat men zich dit wederzijds realiseert. Een essentieel punt is dus, dat er reeds vóór de selectie een duidelijke overeenstemming moet worden bereikt ten aanzien van de criteria die aan de selectie ten grondslag zullen worden gelegd. Uiteraard kunnen hier ethische problemen ontstaan, die echter in deze beschouwingen terzijde worden gelaten. In het volgende zullen wij ervan uitgaan, dat de bedoelde overeenstemming inderdaad te bereiken is.

Het operationeel maken van een criterium

In de praktijk staan we dus dikwijls voor het geval, dat er verschillende en bovendien weinig bewust beleefde criteria aan de selectie ten grondslag dreigen te worden gelegd. Om tot een rationele werkwijze te komen, zullen we eerst moeten trachten, deze criteria expliciet te maken.

Daarmee is onze opgave echter niet ten einde, want op een of andere wijze dient er ook voor te worden gezorgd, dat de criteria toetsbaar worden gemaakt, zij dienen dus een formulering te krijgen, waarin zij zich lenen voor empirisch onderzoek. Dit laatste noemt men het *operationeel maken* van de criteria.

Een relatief nog weinig gecompliceerd voorbeeld kan toelichten, dat wij hier voor een verre van eenvoudige opgave komen te staan. Men denke bijvoorbeeld een afdeling bankwerkerij in een machinefabriek, waar een aantal vacatures (plaatsen) voorkomt voor één functie van constructiemonteur. Men wil deze plaatsen opvullen door uit een groep gegadigden de geschiktste kandidaten te selecteren.

Het bedrijf hanteert primair een economisch gezichtspunt, zodat dus in de eerste plaats wordt gelet op de netto geldopbrengst per werknemer; maar daarnaast stelt men er ook prijs op, dat de sfeer in de afdeling goed is.

Maar deze gezichtspunten blijken in werkelijkheid van complexe aard te zijn: zij omvatten allerlei aspecten, die niet zonder meer van elkaar afhankelijk zijn. Bij nadere analyse vinden we, dat bij de geschiktheidsbeoordeling vooral een rol spelen: de *kwaliteit* van de prestatie — omdat de arbeid een betrekkelijk groot afbreukrisico heeft —; de *kwantiteit* van de prestatie — omdat het rendabel maken van grote investeringen in andere afdelingen mede afhangt van het tempo, waarin de constructiemonteurs hun taak verrichten —; de *mate van zorg* voor gereedschap en werkplaats — zowel met het oog op efficiency en voorkoming van slijtage, als met het oog op het voorkomen van ongevallen —; *toewijding en zelfstandigheid* — het werk vraagt vaak overleg en geduld —; *sociale en disciplinaire aanpassing*.

Het zijn aspecten, die we, bijv. in het kader van 'merit rating', dikwijls tegenkomen. We zullen nu in de eerste plaats voor ieder van hen een maatstaf moeten vinden, om de kandidaten hierop te beoordelen. En als dan deze beoordeling gelukt is, komen we nog voor een volgend probleem te staan: we

zullen de gezichtspunten tegen elkaar moeten afwegen. Want iemand die op een bepaald gezichtspunt boven aan de ranglijst komt te staan, is daarmee nog niet automatisch ook van de topscore ten aanzien van een ander gezichtspunt verzekerd. Als dat laatste wèl het geval zou zijn, zou men uiteraard zich ook niet behoeven in te spannen om alle relevante gezichtspunten op te sporen, men zou met de beoordeling op één gezichtspunt kunnen volstaan.

Hoe stelt men nu een beoordelingsschaal op? Dat is niet altijd een moeilijke taak. Als we bijv. kwantiteit moeten beoordelen van mensen die een rechtstreeks afmeetbare prestatie leveren, kunnen we van deze prestatie uitgaan. Bij metselaars kunnen we vastleggen, hoeveel stenen zij gemiddeld per uur leggen, bij levensverzekeringsinspecteurs in een bepaald rayon, voor hoeveel gulden zij per jaar verzekeringen afsluiten. Ook bij de constructiemonteurs kan men, voor zover zij vergelijkbare arbeid onder vergelijkbare omstandigheden verrichten, een maatstaf vinden in de hoeveelheid gereed product die zij in een tijdseenheid afleveren. Het feit, dat het nodig is een aantal aanvullende kwalificaties te geven t.a.v. arbeid en arbeidsomstandigheden, laat wel zien, dat ook in relatief eenvoudige gevallen het maken van een beoordelingsschaal niet zonder problemen is, maar deze secundaire complicaties laten we verder buiten beschouwing. Belangrijker is, dat er in andere gevallen veel grotere complicaties rijzen. We kiezen nu bijvoorbeeld het gezichtspunt toewijding en zelfstandigheid. Deze beide zijn psychologisch duidelijk verschillend, maar toch zijn zij bij deze functie wel gecorreleerd, omdat de toegewijde werker eerder de gevallen opmerkt, waarin iets bijzonders aan de hand is, waarvoor men zelfstandig een oplossing moet zoeken, terwijl dezelfde toewijding kan voorkomen, dat men te zelfstandig optreedt en dan schade aanricht. Het ligt dus wel voor de hand, dat het bedrijf de beide gezichtspunten samen neemt. Maar bij het opstellen van een beoordelingsschaal ontstaan nu problemen: hoe moet men tegen elkaar afwegen de wèl toegewijde, maar niet zeer zelfstandige man, tegenover de wèl zelfstandige, maar niet uitgesproken toegewijde werker. En hoe plaatst men alle andere denkbare combinaties op de schaal?

In het algemeen kan men zeggen, dat het construeren van een beoordelingsschaal alleen dan betrekkelijk ongecompliceerd is, wanneer de verschillende kwaliteiten die men te beoordelen krijgt ongewrongen in één rangorde van waardering kunnen worden gebracht. In een dergelijk geval spreken we van een *unidimensionele schaal*.

Zuiver unidimensionele schalen treffen we in situaties, waarin we mensen te beoordelen hebben, maar hoogst zelden aan — we zijn gewoonlijk al tevreden met een schaal die bij eerste benadering unidimensioneel is, zoals de bovengenoemde kwantiteitsschaal of — in de psychologie bij het beoordelen van mensen op hun intellectuele potenties — de IQ-schaal, die een indruk geeft van het intellectueel niveau, zonder zich om kwalitatieve intelligentieverschillen te bekommeren.

Bij gezichtspunten zoals toewijding en zelfstandigheid, zorg voor gereedschap en werkplaats, sociale en disciplinaire aanpassing, hebben we telkens met een combinatie van dimensies te maken. Men kan dan toch nog wel een schaal samenstellen, maar dan zal men verschillende onderdelen van een gezichtspunt onderling moeten wegen. Dat moet achteraf toch ook gebeuren, als men uit de verschillende gezichtspunten een totaalbeoordeling wil afleiden. Vaak is het daarom het eenvoudigst, de gezichtspunten zo ver op te splitsen, tot men ieder van hen als unidimensioneel kan behandelen en aan het einde uit al deze schaalwaarden door één weging een totaalbeoordeling af te leiden. De wijze, waarop men uiteindelijk een gezichtspunt operationeel maakt, is sterk afhankelijk van de concrete aard van de betrokken grootheid. Soms beschikt men op enigerlei wijze over een kwantitatieve maatstaf, die aan de praktijk ontleend kan zijn, maar ook eventueel door aflezing van een meetinstrument, bijv. een gekwantificeerde vragenlijst kan worden verkregen. In andere gevallen moet men volstaan met een omschrijving van een aantal karakteristieke gedragsvormen, die men verschillend waardeert: (toewijding) 'hij laat geen produkt uit zijn handen gaan, vóór hij er van overtuigd is, dat het af is', 'als het geheel maar in elkaar zit zonder in het oog lopende fouten, is hij tevreden', 'als het niet direkt past, gaat hij de zaak forceren', etc. In feite houdt dit in, dat men, op een schaal van goed en slecht, een aantal schaalpunten onderscheidt. Minimaal moeten dit er twee zijn, maar zo mogelijk zal men een groter aantal trachten te verkrijgen. Hoe bij de selectie ten aanzien van deze gezichtspunten een beoordeling kan worden verkregen wanneer men de candidaat nog niet in de eigenlijke functie heeft kunnen observeren, is een vraag, die buiten ons onderwerp ligt. Dit is een aangelegenheid, die bijvoorbeeld de psycholoog raakt, die tests construeert, welke geschikt zijn om de schaalbeoordeling ten aanzien van de verschillende gezichtspunten te voorspellen.

Wel van belang voor ons is de wijze, waarop men de verschillende schaalbeoordelingen tot een totaaloordeel moet verenigen. Daartoe moeten we op de hoogte zijn van de validiteit van de verschillende gezichtspunten — of van de gegevens die we gebruiken om de beoordeling op deze gezichtspunten te voorspellen. Kort gezegd komt het hierop neer, dat we veronderstellen, dat het bedrijf achteraf in staat zal zijn de verschillende werkers in dezelfde functie onderling te waarderen, dus een rangorde te maken van meest naar minst waardevol. Indien we nu ook de beoordeling van deze werkers op de verzameling van gezichtspunten kennen, kunnen we trachten door het toekennen van gewichtscijfers deze rangorde zo goed mogelijk uit de gezichtspunten te voorspellen. Als bijvoorbeeld blijkt, dat de bedrijfsbeoordeling veel beter correspondeert met de beoordeling op kwaliteit, toewijding en zelfstandigheid dan met de kwantiteit, zal men aan de eerstgenoemde gezichtspunten een veel groter gewicht toekennen dan aan de laatste — dus geheel analoog aan de wijze waarop ook bij het vaststellen van beloningsnormen via werkclassificatie aan de gezichtspunten via de afweegfactoren een verschillend gewicht

wordt toegekend. De bepaling van deze gewichtscijfers is in feite een kwestie van een statistische techniek: uit de correlaties van de beoordelingen op de gezichtspunten met de totaalwaardering van een groep monteurs, die reeds in de praktijk werkzaam zijn, en uit de correlaties tussen de gezichtspunten onderling, kan men afleiden, welke combinatie van gegevens de totaalwaardering het beste voorspelt. Deze berekening dient dan aan een nieuwe groep werkers geverifieerd te worden, om de invloed van toevalsfactoren zo veel mogelijk uit te schakelen.

Uit het voorgaande volgt wel, dat met het opstellen en operationeel maken van het criterium en met het vaststellen van de wegingsfactoren een omvangrijk vooronderzoek gemoeid is. Het gebeurt maar betrekkelijk zelden, dat een dergelijk onderzoek inderdaad volgens de regels van de kunst wordt uitgevoerd: de hoge kosten en de lange tijd die er mee gemoeid zijn, maken dat dit alleen geschiedt in gevallen, waarin van de selectie zeer veel afhangt. Als met de opleiding bijvoorbeeld van een straaljagerpiloot bedragen gemoeid zijn die in de honderdduizenden lopen, loont ook een geringe verbetering van de selectieprocedure reeds, zelfs als deze slechts na een kostbaar onderzoek wordt bereikt.

In de praktijk ziet men dikwijls dat volstaan wordt met betrekkelijk grove schattingen, die vaak gebaseerd zijn op de intuïtie, eventueel aangevuld met enkele kleine verkennende onderzoeken. Dergelijke schattingen behoeven niet ondeugdelijk te zijn, maar zij moeten uiteraard achteraf wel terdege op hun geldigheid worden getoetst.

Betreft het een functie, waarin maar een klein aantal plaatsen beschikbaar is, dan is zelfs een verkennend onderzoek gewoonlijk niet mogelijk, zodat men geheel op het kompas van de intuïtie moet varen. Wat in ieder geval zou moeten geschieden — maar waar in de praktijk, onder meer om financiële redenen, gewoonlijk ook veel te weinig van terecht komt — is de verificatie achteraf van de gekozen criteria en bijbehorende selectiemiddelen. Deze verificatie wordt gewoonlijk als validatie betiteld.

Validatie

Het probleem van de validering van een selectieprocedure is verre van eenvoudig. Op de technische aspecten van dit probleem gaan wij niet in, maar wel willen wij op een aantal principiële punten de aandacht vestigen. Het eerste is, dat een dergelijke validering uiteraard alleen zin heeft, indien aan de selectie en de beoordeling dezelfde gezichtspunten ten grondslag worden gelegd. Deze uitspraak schijnt een vanzelfsprekende waarheid, maar in werkelijkheid wordt dikwijls tegen deze regel gezondigd. Het is bijvoorbeeld heel goed mogelijk, dat bij het formuleren van de selectiecriteria de zwaarste nadruk wordt gelegd op de kwaliteitseis, terwijl achteraf de beoordeling primair op grond van de kwantiteit van de producties geschiedt.

Wanneer selectie en beoordeling door verschillende personen geschieden, die bovendien de criteria waarvan zij uitgaan weinig expliciet hebben gemaakt, kunnen dergelijke tegenstellingen zeer gemakkelijk voorkomen. Met name in industriële situaties waar bij de validatie van selectieprocedures gewoonlijk wordt uitgegaan van beoordelingen die worden gegeven door bazen en bedrijfsleiders, stuit men zeer dikwijls op problemen die kennelijk voortkomen uit het hanteren van andere criteria dan bij de selectie geschiedde.

Een tweede punt is de mate van de betrouwbaarheid van de beoordelingsprocedure. Talrijke onderzoeken hebben aangetoond, dat de mate van overeenstemming van competent te achten beoordelaars, aangaande dezelfde persoon of hetzelfde geval, gewoonlijk veel te wensen overlaat. De nogal eens toegepaste praktijk om door een 'conferentiemethode' verschillen in beoordeling te doen verdwijnen, voert dikwijls tot schijn-overeenstemmingen. Nu spreekt het wel vanzelf, dat de mate van betrouwbaarheid der beoordeling een grens stelt aan de validiteit van de selectieprocedure. Ook bij een ideale selectiemethode blijven er verschillen tussen voorspelling en praktijkbeoordeling, die voortkomen uit de toevallige variaties tussen de beoordelaars of ook uit de toevallige variatie binnen één beoordelaar die hetzelfde gedrag de ene keer anders waardeert dan de volgende maal. Merkwaardigerwijze geeft men zich in de praktijk zelden rekenschap van de invloed die deze inconsistentie van een beoordelaar of tussen beoordelaars heeft: als eenmaal een eindoordeel over iemand is gegeven, handelt men verder alsof alle verschillen tussen selectie en beoordeling voortkomen uit onvolmaaktheden van de selectieprocedure.

In werkelijkheid zien we dus, dat een geringe mate van overeenstemming tussen de waardering op grond van de selectie en die op grond van de beoordeling kan voortkomen uit onvolkomenheid van de beoordeling, of uit tekorten in de selectieprocedure, of ook uit verschillen in de criteria die aan beoordeling en selectie ten grondslag worden gelegd, terwijl uiteraard ook nog externe foutenbronnen bestaan, bijvoorbeeld veranderingen in de persoon of de situatie van de geselecteerde.

HOOFDSTUK II

PREDICTOREN

Inleiding

In het eerste hoofdstuk werden enkele grondbegrippen besproken, die voor de selectieproblematiek van belang zijn. Thans zullen wij nader ingaan op het vraagstuk van de *predictoren*. Bij selectie immers beoordelen wij mensen en situaties ten aanzien van de waarschijnlijke ontwikkeling van een gegeven persoon in een gegeven situatie. Deze ontwikkeling is op het tijdstip van selectie nog niet gegeven, voorlopig is er zelfs alleen nog maar sprake van de mogelijkheid dat deze mens in deze situatie zal worden geplaatst. Over de wijze waarop hij zich zal ontplooiën als deze mogelijkheid tot werkelijkheid wordt, kunnen wij alleen maar veronderstellingen maken. En wanneer wij aan deze veronderstellingen werkelijkheidskarakter toekennen, kunnen wij ze betitelen als voorspellingen. Voorspellingen zullen evenwel op iets gebaseerd moeten zijn, namelijk op bepaalde gegevens waarvan gebleken is (of eventueel verondersteld wordt), dat zij gecorreleerd zijn met het criterium waarop achteraf de beoordeling van de juistheid der selectie geschiedt. De grootheden waarop deze gegevens betrekking hebben noemen we *predictoren*.

De keuze van de predictoren

Voor al in situaties waarin tal van mogelijkheden aanwezig zijn ten aanzien van de criteria die aan de selectie ten grondslag kunnen worden gelegd, kan de keuze van geschikte predictoren een moeilijk probleem opleveren. Ten dele wordt deze keuze veelal intuïtief gemaakt. Wanneer men bijvoorbeeld besluit tot een psychologisch of medisch onderzoek in verband met selectie, houdt dit in, dat er predictoren zullen worden gebruikt die ontleend zijn aan het psychologisch of medisch instrumentarium. Degene die dit besluit neemt veronderstelt daarmee impliciet, dat dergelijke predictoren voor het gegeven geval vermoedelijk validiteit zullen bezitten. Dit behoeft lang niet altijd bewezen te zijn: er zijn gevallen, waarin bijvoorbeeld een bedrijf iedere sollicitant die de zeef van een bepaalde formele voorselectie gepasseerd is naar de bedrijfsarts of naar een psychologisch instituut zendt; zo schrijft ook de huidige Rijkssubsidieregeling voor de beroepskeuzevoorlichting voor, dat in elk adviesgeval een psychologisch onderzoek moet worden ingesteld. In andere selectiesituaties zal wellicht niet op zo automatische wijze worden gehandeld, maar zal men op grond van bepaalde overwegingen voor een concreet geval tot de conclusie komen, dat bij de selectie gebruik zal worden gemaakt van predictoren van specifieke, bijvoorbeeld psychologische aard. De keuze van deze predictoren is dan vervolgens een opgave die aan de psycholoog — of in andere gevallen aan de arts, de personeelchef of de bedrijfsveiligheidsdienst — moet worden overgelaten.

Op dit punt in het keuzeproces is het noodzakelijk, eerst de beoordelingscriteria expliciet te maken, indien dit althans op eenvoudige wijze — onder andere door het voeren van gesprekken met mensen die het resultaat van de selectie moeten beoordelen — mogelijk is. Zijn de criteria bekend, dan kan men veelal teruggrijpen op hulpmiddelen, bijvoorbeeld tests, waarvan uit de ervaring reeds bekend is, dat deze in verband met deze criteria een zekere mate van validiteit bezitten. Zijn zij niet beschikbaar, dan zal men de hulpmiddelen zelf moeten construeren, bijvoorbeeld door een vereenvoudigd model te maken van de situatie waarin de betrokkene later eventueel zal worden geplaatst, om te kunnen observeren hoe de betrokkene zich daarin gedraagt onder bepaalde experimentele condities. In andere gevallen zal men trachten door 'mentale analyse' de kwaliteiten op te sporen die in verband met een gegeven criterium een rol zouden kunnen spelen, om vervolgens weer de instrumenten te kiezen die voor het onderzoek van deze kwaliteiten geschikt zijn.

Er zijn ook gevallen, waarin het niet goed gelukt om op korte termijn de beoordelingscriteria expliciet te maken; dit geldt zeker voor de meeste onderwijssituaties, maar ook nogal eens bij selectie voor bedrijfsfuncties. Wil men dan toch selecteren, dan zal de onderzoeker een intuïtieve keuze moeten maken, hij kan dan bijvoorbeeld van mening zijn, dat een intelligentietest voor de voorspelling van onderwijsresultaten van waarde zal zijn of dat een onderzoek naar het concentratievermogen valide inlichtingen zal kunnen geven over de wijze waarop iemands geschiktheid als chauffeur t.z.t. door een gegeven bedrijf zal worden beoordeeld.

Verificatie van de keuze

Hoewel men dus bij deze keuze van predictoren zoveel mogelijk uitgaat van het gezichtspunt van de validiteit, moet de werkelijke waarde van de gemaakte keuze nog blijken. Wil men wat dit betreft verantwoord werken, dan kan men kiezen tussen twee mogelijkheden.

De eerste is een onderzoek, met behulp van de voorlopig gekozen predictoren, van een groep personen, bij wie reeds een beoordeling op grond van het criterium-achteraf mogelijk is, waarbij we in het midden laten hoe deze beoordeling dient te geschieden opdat zij zo betrouwbaar mogelijk is. Wil men bijvoorbeeld constructiebankwerkers selecteren, dan zou men kunnen uitgaan van een groep bankwerkers die reeds in het bedrijf werkzaam zijn en waarvan een beoordeling heeft plaatsgevonden. Liefst zal men dan werkers van uiteenlopende kwaliteit kiezen, goeden, middelmatigen en slechten (niet uitsluitend goeden en slechten, omdat dan niet zonder meer duidelijk wordt, in welke opzichten de middelmaat zich van deze beide uitersten onderscheidt). Deze groep wordt nu onderzocht met de voorlopig gekozen predictoren en men gaat na, welke combinatie van gegevens het praktijkoordeel het beste benadert.

Hoewel deze methode in de praktijk wel wordt toegepast, heeft zij toch een belangrijk bezwaar: een dergelijke proefgroep is niet geheel vergelijkbaar met de groep personen die men in werkelijkheid voor selectie moet onderzoeken. De proefgroep bestaat uit mensen die reeds een vrij belangrijke selectiegang hebben doorgemaakt: zij is geselecteerd vóór de aanstelling en ook in de praktijk zullen velen der zwakken nog zijn afgevallen, eventueel hebben goede werkers door promotie de groep eveneens verlaten. Bovendien is de proefgroep gewoonlijk naar leeftijd anders samengesteld dan de eigenlijke onderzoekgroep. Zelfs kleine leeftijdsverschillen kunnen hier soms zeer belangrijk worden, onder meer bij onderwijsselectie. Tenslotte speelt ook de scholing een rol, de proefgroep beschikt over een praktijkervaring en soms ook over extra-kennis, die bij de eigenlijke onderzoekgroep niet aanwezig is, terwijl het ook kan voorkomen, dat door de praktijk eigenschappen tot ontwikkeling worden gebracht, die bij kandidaten voor een functie nog niet in die mate aanwezig zijn. Het vooronderzoek geeft dus wel aanwijzingen over de validiteit van de voorgestelde predictoren, maar geen definitief uitsluitsel; men moet toch de resultaten van de selectiegang in de praktijk afwachten.

De tweede mogelijkheid is, met de voorlopig gekozen predictoren een onderzoek uit te voeren op de categorie personen waarbij de selectie moet worden uitgevoerd, zonder aan de uitkomsten van dit onderzoek consequenties te verbinden ten aanzien van de selectie. De resultaten van het onderzoek worden dan bij voorkeur geheim gehouden voor degene die de keuze moet doen, om onbewuste beïnvloeding te voorkomen; zij worden echter wel geregistreerd. Na verloop van tijd, als de beoordeling op het criterium-achteraf bekend is geworden, kan men nu nagaan, in hoeverre de predictoren een voorspelling van deze beoordeling mogelijk zouden hebben gemaakt.

Liefst moet men dan de categorie onderzochten nog in twee groepen verdelen, waarbij onafhankelijk van elkaar wordt nagegaan of voorspelling mogelijk is; dit om de kans op toevallige overeenstemming tussen de onderzoeksgegevens en de beoordeling te verminderen.

Een dergelijke werkwijze is echter alleen mogelijk in gevallen waarin de selectie betrekking heeft op grote groepen personen en bovendien ook periodiek moet geschieden. Geschiedt de selectie niet periodiek, dan heeft het vooronderzoek geen praktische zin, terwijl bij een klein materiaal de kosten van het vooronderzoek bezwaarlijk zullen zijn. Verder doet deze methode uiteraard nogal een beroep op het geduld van de belanghebbenden.

Voor onderwijsselectie, voor toepassing van psychologische selectiemethoden, bijvoorbeeld bij recrutenonderzoek, maar ook bij geregeld terugkerende onderzoekingen van groepen personen voor specifieke functies in grotere bedrijven, is deze tweede methode echter de beste — hetgeen niet wil zeggen, dat zij daar in de praktijk ook steeds zou worden gebruikt!

Eisen, aan de predictoren te stellen

In het voorgaande zijn het onderzoekinstrument en de predictor niet zeer scherp van elkaar onderscheiden. In vele gevallen zal een onderzoekinstru-

ment, bijvoorbeeld een test, slechts één informatie leveren die als predictor kan dienen. Het kan echter ook zo zijn, dat één instrument verschillende predictoren levert. Een intelligentietest kan bijvoorbeeld, naast een aanwijzing over het intellectueel niveau in het algemeen, ook informatie geven over deelaspecten van de intelligentie — onder andere het denken in woorden, het denken in getalbegrippen, het denken in ruimtelijke relaties —, terwijl eventueel het aantal gemaakte fouten nog weer andere aanwijzingen geeft. Ieder van deze gegevens kan als een afzonderlijke predictor worden gebruikt. Aan een bruikbare predictor stellen we in de eerste plaats de eis, dat deze kwantitatief volgens één dimensie kan worden vastgelegd. Als we dus een groep personen met het betreffende instrument onderzoeken, moeten we de gegevens van de predictor in kwestie kunnen rangschikken in één rangorde van hoog naar laag. Anders gezegd: als twee personen, A en B, ten aanzien van deze predictor vergeleken worden, mag dit niet tot de slotsom voeren, dat in bepaalde opzichten A boven B en in andere B boven A staat. Is dit toch het geval, dan zal men moeten trachten de informatie op te splitsen in een aantal kleinere eenheden, die wel als unidimensioneel kunnen worden behandeld — of we moeten op één of andere wijze een combinatieregel vinden, waardoor we de tegenstrijdige informatie toch tot één totaalbeoordeling kunnen samenvatten. Reeds eerder is gezegd, dat de eis van unidimensionaliteit in de praktijk niet al te stringent kan worden gehanteerd, gewoonlijk zullen we de realiteit wel enig geweld moeten aandoen, om met ééndimensionele beoordelingen te kunnen werken.

Om een kwantitatieve beoordeling mogelijk te maken, moeten we ten aanzien van de predictorgegevens ten minste twee categorieën kunnen onderscheiden: 'ja — nee' of 'hoog — laag'. Deze categorieën zullen soms scherp kunnen worden onderscheiden, bijvoorbeeld 'man — vrouw' of 'wel — niet een bepaald schooltype bezocht', maar in andere gevallen vloeiend in elkaar overgaan, bijvoorbeeld 'introvert — extravert', of 'wel — niet ijverig'.

Bij vloeiende overgangen zullen we dikwijls in staat zijn meer dan twee categorieën te onderscheiden en dus verschillende gradaties te maken.

Wordt dat aantal gradaties zeer groot, dan krijgen we tenslotte de continue schaal waarop de beoordeling plaats kan vinden (lengte, gewicht, temperatuur, intelligentiequotient en dergelijke). Voor de verfijning van de predictie is het aantal gradaties natuurlijk wel van belang, maar voor het principe van de selectie maakt het weinig uit.

Van veel belang is daarentegen een aantal andere kenmerken van de predictor, in de eerste plaats de *betrouwbaarheid*. Hieronder verstaan we, dat een herhaling van de meting ons over de predictor dezelfde informatie zou moeten geven als bij het eerste onderzoek werd verkregen. Als we bijvoorbeeld tweemaal de lichaamstemperatuur meten, zouden beide metingen tot dezelfde uitkomst moeten leiden.

Het voorbeeld laat zien, dat hier problemen kunnen optreden. De temperatuur op verschillende plaatsen van het lichaam gemeten, is niet precies dezelf-

de; twee thermometers kunnen op dezelfde plaats en tijd verschillende temperatuurwaarden aangeven en bovendien kan tussen twee metingen de temperatuur zich wijzigen. Dit laatste aspect, dat als het ware betrekking heeft op de 'betrouwbaarheid' van de onderzochte persoon, duiden we aan met de term *stabiliteit*. Vooral bij het onderzoek van jeugdige personen stuiten we nogal eens op kenmerken van geringe stabiliteit. Ongeacht de deugdelijkheid van het onderzoekinstrument lenen dergelijke kwaliteiten zich in het algemeen weinig als grondslag voor het doen van predicties, omdat het toevallige ogenblik van onderzoek teveel invloed heeft op de uitkomsten. Verder zullen we duidelijk moeten vastleggen op wat voor wijze en met welke hulpmiddelen de meting van de predictor wordt uitgevoerd, om ook deze foutenbron uit te schakelen.

Pas daarna heeft het zin, aandacht te geven aan de betrouwbaarheid van het onderzoekinstrument, speciaal ten aanzien van het aspect dat we als predictor willen gebruiken (het is immers ook nog mogelijk, dat één instrument ten aanzien van verschillende aspecten niet op gelijke wijze betrouwbaar is). In het bijzonder bij psychologisch onderzoek stuit men dan op de moeilijkheid, dat de eerste meting een zeker oefeningseffect heeft, dat van invloed is op de uitkomst van een tweede meting. Naarmate men meer tijd tussen de beide metingen laat verlopen, zal dit oefeningseffect in de regel kleiner worden, maar dan neemt weer de kans toe, dat de onderzochte persoon intussen verandert. Op verschillende wijzen heeft men in de psychologie toch oplossingen gevonden, waardoor het desondanks mogelijk is de betrouwbaarheid van de predictoren vast te leggen en men heeft ook minimumeisen geformuleerd waaraan de instrumenten in dit opzicht moeten voldoen.

Een bijzonder punt is daarbij nog, dat de betrouwbaarheid van een instrument niet voor alle toepassingen dezelfde behoeft te zijn, er moet dus op worden gelet, dat ook aan de eisen voor de concrete toepassing wordt voldaan.

Hierbij is vooral van belang, dat het instrument in de categorie te onderzoeken personen voldoende differentieert. Een intelligentietest die voor meer dan middelmatig begaafde volwassenen ontwikkeld is kan voor toepassing aan het einde van de lagere school weinig geschikt zijn, omdat de meeste opgaven voor vrijwel alle kinderen te moeilijk zijn en er dus maar weinig spreiding in de resultaten is. Door deze geringe spreiding wordt niet alleen de praktische bruikbaarheid van de testgegevens kleiner, maar vermindert gewoonlijk ook de betrouwbaarheid.

Naast de betrouwbaarheid zullen gewoonlijk ook allerlei praktische overwegingen invloed hebben op de keuze van de predictoren. Met name kan hier worden gedacht aan het economisch aspect: sommige gegevens zijn op veel eenvoudiger en goedkoper wijze te verkrijgen dan andere. Heeft men de keuze tussen twee even valide informatiebronnen, dan zal men allicht degene prefereren, die economisch het meest verantwoord is. Om deze reden wordt bijvoorbeeld bij psychologische selectie veelvuldig gebruik gemaakt van onderzoeksmethoden die collectief kunnen worden toegepast — zoals schriftelijke

tests. Doordat hier de tijd die de onderzoeker per individu moet besteden gering is, speelt de tijdsduur van een collectieve test ook geen grote rol en door een test betrekkelijk lang te maken kan men de betrouwbaarheid van de uitkomsten gunstig beïnvloeden. Daardoor kan eventueel de economisch meest verantwoorde methode tevens de meest betrouwbare zijn!

De kansverwachting van de geschiktheid

Wanneer we beschikken over een valide predictor zouden we uiteraard gaarne zien, dat deze ons in staat stelt een zeer scherpe scheiding te maken tussen geschikten en ongeschikten. Idealiter gesproken: wanneer we weten, welke niveaueisen in de functie worden gesteld, zouden we wensen, dat we een 'critisch punt' op de predictor zouden kunnen aangeven, zodanig, dat vrijwel ieder die tenminste dit punt bereikt geschikt zal blijken te zijn en vrijwel iedereen die het niet bereikt ongeschikt. In verband met onderwijsselectie is hier door KOHNSTAMM de term 'loketkastmethode' voorgesteld. Zouden we de prestaties — die we voortaan *scores* zullen noemen — op een predictor die aan deze ideaalvoorwaarde voldoet op hun frequentie bezien, dan zouden we vinden, dat de frequentieverdeling twee toppen heeft: één voor de scores van de ongeschikten en één voor de scores van de geschikten, met tussen deze toppen een dal, dat liefst zo diep mogelijk zou moeten zijn. Immers, in dat dal kunnen we het kritische punt leggen, waar de scheidingslijn tussen ongeschikt en geschikt moet worden getrokken; hoe dieper het dal, hoe geringer het aantal fouten dat we maken.

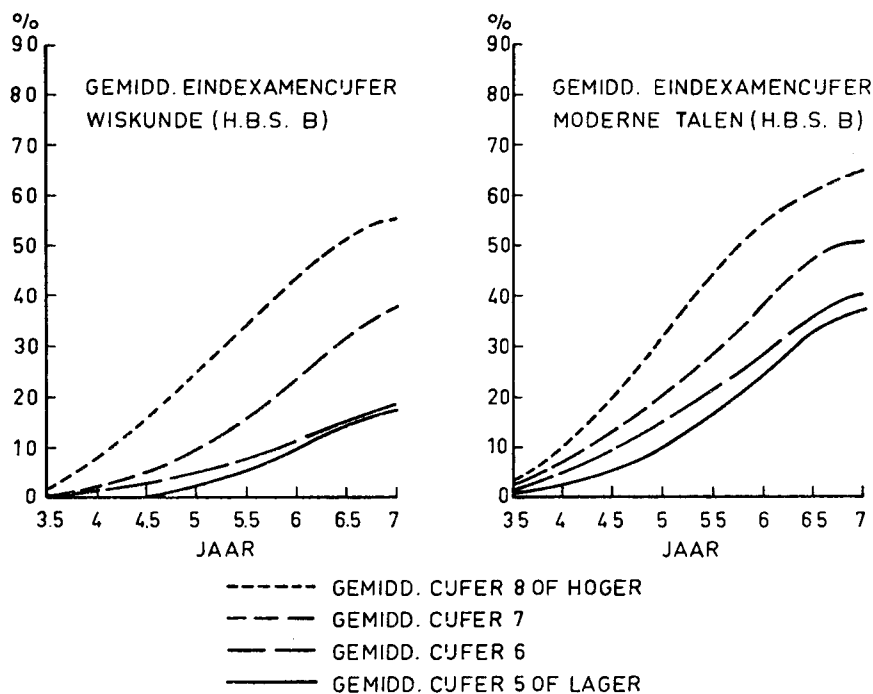
Helaas beantwoordt de realiteit niet aan dit ideaalbeeld. In werkelijkheid is er een vloeiende overgang tussen geschikt en ongeschikt, ook al moet men ergens een grenslijn trekken. Bij individuen die een marginale positie innemen zal het dikwijls van kleine toevallige omstandigheden en vaak ook van de willekeur van de beoordelaar afhangen, aan welke zijde van de grenslijn zij belanden. Daardoor is er nooit één punt, waar een scherpe scheiding tussen geschikten en ongeschikten kan worden gemaakt, maar men vindt, dat met het hoger worden van de score het percentage geschikten toeneemt of — van het individu uit gezien — de kans van slagen groter wordt. De relatie tussen predictorscore en beoordeling achteraf kan altijd slechts worden gegeven in termen van een dergelijke *kansverwachting*.

In het volgende zullen wij ter vereenvoudiging van de discussie de beoordeling-achteraf beperken tot een onderscheiding van geschikt en ongeschikt. In principe geldt echter bij een verfijnder beoordeling geheel dezelfde gedachtingang als hierna wordt weergegeven.

Ter toelichting van de aard van de kansrelatie willen we thans een aantal voorbeelden uit de praktijk geven. Het eerste heeft betrekking op de voorspelling van het succes in een academische studie, wanneer we beschikken over bepaalde gegevens betreffende de eindexamenresultaten. De predictor is hier dus een eindexamenvak, de score het eindexamencijfer, terwijl als criterium voor de beoordeling achteraf het al of niet afleggen van het candi-

daatsexamen wordt gekozen. Over dit onderwerp heeft het Centraal Bureau voor de Statistiek onderzoeken verricht, die in 1958 zijn gepubliceerd.*) De hierna volgende afbeeldingen zijn bewerkt naar enkele grafische voorstellingen uit deze publicatie.

Voor de technische wetenschap, waaronder hier moet worden verstaan het geheel van de studies te Delft, is nagegaan de relatie tussen het gemiddeld cijfer voor de moderne talen en het al of niet afleggen van het kandidaatsexamen binnen een bepaald aantal jaren. Ook de relatie tussen het gemiddelde cijfer voor de wiskundevakken op het eindexamen en het slagen voor het kandidaatsexamen is onderzocht. In beide gevallen is het eindexamen in kwestie dat van de h.b.s.-B. Afbeelding 1 geeft de relatie tussen eindexamencijfer en slagen voor het kandidaatsexamen weer. De grafieken hebben betrekking op de studieresultaten binnen zeven jaar na het begin van de studie. Doordat het kandidaatsexamen in de technische studie gewoonlijk kort voor het afstuderen valt, is het percentage studenten, dat na meer dan zeven jaar dit examen aflegt, zeker niet te verwaarlozen. Het verloop van de curven geeft echter wel een aanduiding, waar ongeveer de bovengrens zal liggen. Ter toelichting in de eerste plaats een opmerking over het feit, dat de percentages geslaagden bij de talencijfers veel hoger liggen dan bij de wiskunde-



Afbeelding 1.

*) Een literatuuroverzicht wordt aan het einde van hoofdstuk V gegeven.

cijfers, terwijl de gegevens toch betrekking hebben op dezelfde personen, dwz.: het gemiddelde percentage geslaagden moet uiteraard bij de talen even hoog zijn als bij de wiskunde. Hieruit volgt weer, dat het gemiddelde wiskundecijfer bij de technische studenten hoger moet zijn geweest dan het gemiddelde talencijfer — een conclusie, die wel aannemelijk mag worden geacht. De vraag blijft dan nog, waarom een hoog gemiddeld talencijfer een betere predictor is dan een hoog wiskundecijfer. Men kan vermoeden, dat dit komt omdat technische studenten met hoog gemiddeld talencijfer als regel ook een hoog of vrij hoog gemiddeld wiskundecijfer zullen hebben — dit gezien de aard van de technische studie — terwijl het omgekeerde in veel mindere mate geldt. Dit zou te verifiëren zijn door de predictieve werking van de talencijfers te analyseren bij constant gehouden wiskundecijfers, maar hierover zijn nog geen onderzoeken gepubliceerd.

We zien dan voorts, dat steeds geldt, dat bij een hoog cijfer een betere kans van slagen behoort. Maar het is niet zo, dat degenen die een hoog gemiddeld cijfer hebben vrijwel zeker zullen slagen en dat degenen met een laag gemiddeld cijfer vrijwel zeker zullen mislukken. Zelfs met een zes of lager als gemiddeld wiskundecijfer ligt de kans dat men t.z.t. een kandidaatsexamen in de technische wetenschappen zal afleggen, *als men besloten heeft deze studie te kiezen*, nog in de orde van 20% en een slagen voor het kandidaatsexamen houdt in dit geval vrijwel zeker in, dat men ook het ingenieursexamen zal afleggen. De gecursiveerde tussenzin wijst er op, dat de bevinding niet mag worden gegeneraliseerd voor een willekeurig eindexaminandus met een gemiddelde van zes of minder voor de wiskundevakken! Uiteraard is 20% geen grote kans op succes, wanneer het gaat om een keuze waar voor de toekomst zeer veel van afhangt en men kan er over twisten, of men dus iemand met deze wiskundecijfers de studie niet ten sterkste zou moeten ont-raden*), maar geheel te verwaarlozen is de kans op succes toch ook niet.

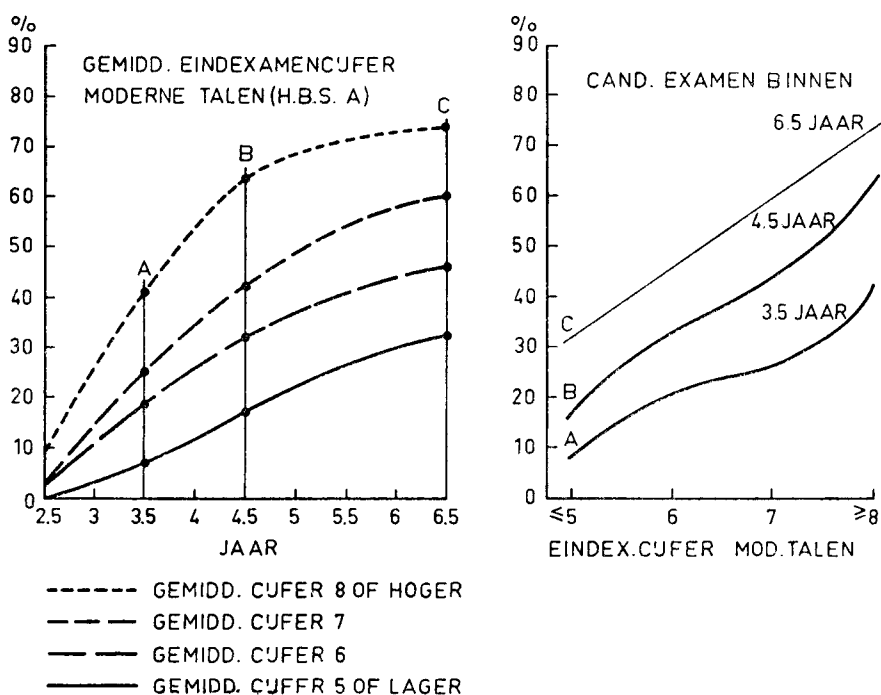
Analoge kansrelaties heeft men ook gevonden voor het verband tussen eind-examencijfers en het behalen van het kandidaatsexamen in andere studierichtingen; vrijwel steeds geldt, dat hoe hoger het gemiddelde examencijfer, des te groter ook de kans van slagen. Een opvallende uitzondering vormt de rechtenstudie, waar zowel voor het gemiddeld cijfer voor de moderne talen (gymnasium A en B) als voor het gemiddeld wiskundecijfer (gymnasium B) slechts een zeer geringe relatie tussen de hoogte van het cijfer en de kans van slagen werd gevonden. Mede gezien het feit, dat de totale kans van slagen bij de rechtenstudie hoog ligt (in de orde van grootte van 80%) kan men zeggen, dat de eindexamencijfers althans voor de moderne talen en wiskunde hier vrijwel geen predictieve waarde hebben voor de voorspelling van het studiesucces.

*) De gegevens maken geen onderscheid tussen de verschillende studierichtingen in Delft; het is echter a priori niet onwaarschijnlijk, dat er tussen de vakken onderling wel verschillen in prognose zullen bestaan.

De kansrelatie kan men ook op een andere wijze benaderen, hetgeen in afbeelding 2 wordt gedemonstreerd. Uitgangspunt is een grafiek analoog aan die voor de technische wetenschap, deze is eveneens naar de bedoelde C.B.S.-publicatie bewerkt. De gegevens betreffen ditmaal de predictieve waarde van het eindexamencijfer voor de moderne talen op h.b.s.-A, bij de voorspelling van studiesucces in de opleiding voor econometist.

Men ziet ook hier weer het beeld van een zeer duidelijk verschil in studiesucces bij de verschillende examencijfers, maar opnieuw geen indicatie van een benaderen van de absolute zekerheid ten aanzien van slagen of falen.

Nu kan men op verschillende punten van de grafiek een verticale lijn trekken. De snijpunten van deze lijnen met de predictiecurven geven dan de percentages dergenen, die binnen een zeker aantal jaren het candidaatsexamen hebben behaald. Deze percentages kunnen we overbrengen naar een nieuwe grafiek, waarbij thans in horizontale richting het eindexamencijfer is uitgezet. Door de punten die bij één snijlijn behoren onderling te verbinden krijgen we thans een aantal nieuwe kansverwachtingslijnen, die ons nu rechtstreeks een beeld geven van de relatie tussen de predictor en het criterium. Men ziet dat het beeld van de relatie verandert naarmate de termijn waarop men voorspellingen doet langer wordt. Het beeld van deze relatie verschilt ook met de con-



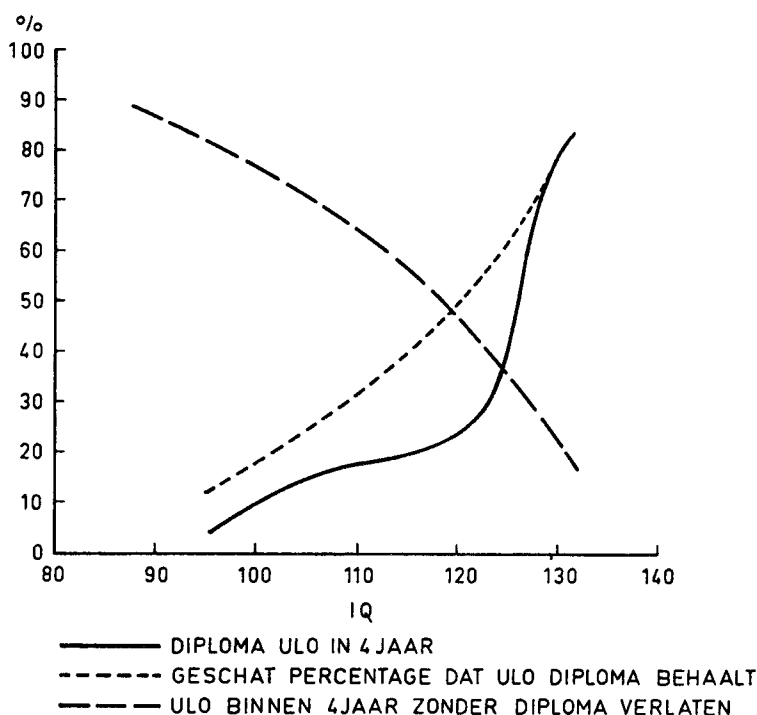
Afbeelding 2.

crete studie waarvoor men voorspellingen doet, zoals men kan constateren wanneer men zelf uit afbeelding 1 soortgelijke kansverwachtingslijnen afleidt voor de technische studie.

Wat hier gevonden is voor de waarden van schoolvorderingsgegevens als predictor, blijkt ook voor gegevens van andere aard op te gaan. In afbeelding 3 geeft de getrokken lijn de relatie aan tussen IQ en kans van slagen — zonder doubleren — voor het m.u.l.o.-examen. De gegevens zijn ontleend aan een onderzoek van prof. Nieuwenhuis op Haagse u.l.o.-scholen.

In de afbeelding is ook aangegeven de kansverwachting ten aanzien van het voortijdig verlaten van de school, gemeten over een periode van vier jaar. Sommeert men de percentages van de gediplomeerden en de uitvallers, dan komt men niet op 100%, omdat een gedeelte der leerlingen één of meermalen niet bevorderd is en zich nog op school bevindt. Uit de afbeelding blijkt, dat dit vooral leerlingen met een IQ uit het middengebied betreft — zoals bij een valide predictor ook mag worden verwacht.

Als men er van uitgaat, dat de meeste leerlingen die zich na 4 jaar nog op school bevinden t.z.t. wel een diploma zullen behalen, kan men uit de gegevens ook een schatting maken van de kansverwachting ten aanzien van het behalen met of zonder dubblures van een m.u.l.o.-diploma, afhankelijk van



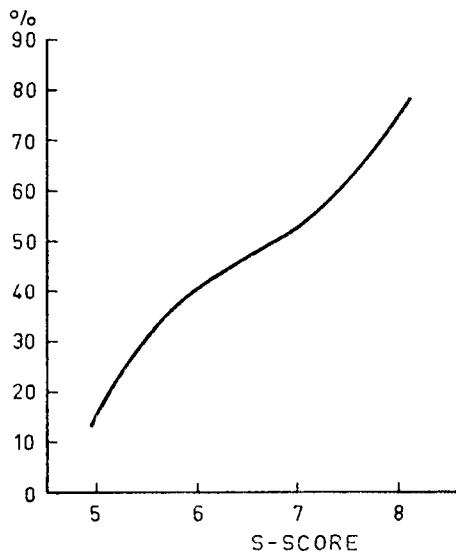
Afbeelding 3.

het IQ dat bepaald werd bij de intrede in de school. Dit geschatte verband is in de figuur door een gestippelde lijn aangegeven. Bij de interpretatie dient nog bedacht te worden, dat leerlingen met IQ hoger dan 125 in het u.l.o. slechts met betrekkelijk geringe frequentie voorkomen, zodat de schatting in hun geval maar zeer globaal kan zijn.

Analoge relaties blijken ook te bestaan tussen het IQ en de kans van slagen in andere vormen van voortgezet onderwijs. In het volgende hoofdstuk zal nog de relatie worden besproken die er bestaat tussen het IQ van kandidaten voor het toelatingsexamen v.h.m.o. en de kans op het behalen van het einddiploma. In alle gesproken gevallen geldt, dat de weergegeven percentages afkomstig zijn uit eerste onderzoeken, zodat het precieze verloop van de kansverwachtingscurven bij nader onderzoek wel enige wijziging zal kunnen ondergaan. Het principe van de relatie staat echter wel vast.

Over het verband tussen predictorscores en functiegeschiktheid in het bedrijfsleven zijn in ons land nog nauwelijks gegevens gepubliceerd. Men mag echter wel aannemen, dat zowel principieel als praktisch hier geheel analoge relaties bestaan, gezien het feit, dat dergelijke verbanden bij Amerikaanse onderzoeken wel zijn gevonden. Zij gelden bovendien niet alleen voor psychologische kenmerken en schoolvorderingsgegevens, maar ook voor predictoren van andere, bijvoorbeeld sociologische of somatische aard.

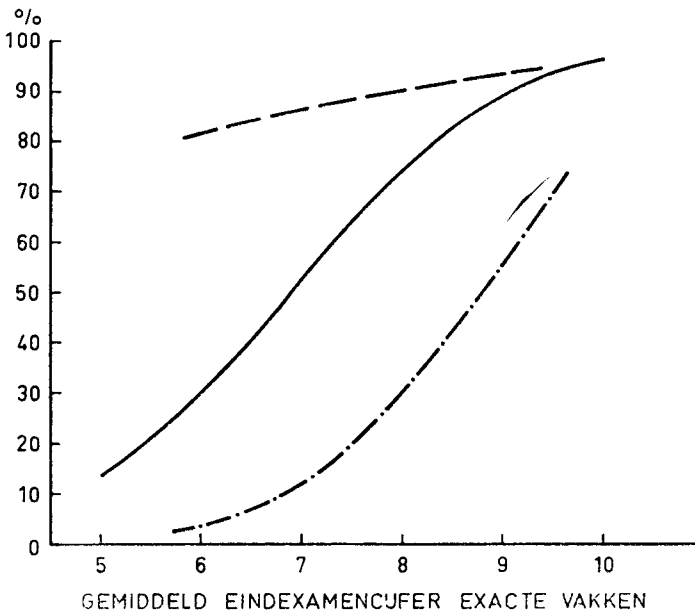
Het verband blijft ook bestaan, wanneer men gegevens van verschillende predictoren gaat combineren. Neemt men bijvoorbeeld niet de resultaten voor één specifiek onderwijsvak of groep van vakken als uitgangspunt, maar een gemiddelde, dat de prestaties in het onderwijs als geheel representeert, dan zal men evenmin als in de voorgaande gevallen bereiken, dat er een twee-



Afbeelding 4.

toppige curve ontstaat, die geschikten van ongeschikten onderscheidt. En hetzelfde geldt, als men bijvoorbeeld schoolvorderingengegevens combineert met een gegeven over de werkhouding. Van deze beide gevallen geven de afbeeldingen 4 en 5 een indruk.

Afbeelding 4 is afgeleid uit een onderzoek van MATTHIJSEN en SONNEMANS, dat onder meer betrekking had op het studierendement van het v.h.m.o., gezien vanuit de prestaties aan het einde van de lagere school. Deze prestaties zijn gemeten met behulp van een serie van 18 schoolvorderingentests, afgenomen in een representatieve steekproef van zesde klassen van lagere scholen in Noord-Brabant. Van de leerlingen die tot het v.h.m.o. werden toegelaten, is nagegaan, welk percentage na $4\frac{1}{2}$ jaar nog dit onderwijs bezocht en bovendien niet meer dan één jaar vertraging had. De beoordeling bij de schoolvorderingentests had plaats volgens een schaal van 1–10, waarbij het cijfer 5 echter niet als gebruikelijk 'juist onvoldoende' betekende, maar integendeel de gemiddelde prestatie representerende.



Afbeelding 5. (verklaring in de tekst).

Afbeelding 5 heeft betrekking op een onderzoek ingesteld op de Technische Hogeschool te Delft. Het gaat bij deze afbeelding om het percentage studenten, dat het eerste examen (P1) 'op tijd', dwz. na één jaar, met goed gevolg heeft afgelegd. Dit percentage is eerst gezien afhankelijk van het gemiddelde eindexamencijfer voor de exacte vakken (de getrokken lijn in de figuur) en vervolgens afhankelijk van dit eindexamengemiddelde plus een beoordeling van de vorderingen en 'ijver', zoals deze enige tijd na het begin van de studie

door de instructeurs van de wiskundepractica werd gegeven. Weergegeven worden de kansverwachting voor studenten met ruim voldoende of goede vorderingen (streeplijn) en voor studenten met voldoende of slechte vorderingen en bovendien onvoldoende ijver (streep-puntlijn). De resultaten voor de overige studenten liggen tussen deze beide lijnen.

Op de vraag, op welke wijze op grond van dergelijke kansverwachtingsgegevens op verantwoorde wijze geselecteerd kan worden, zal in een volgende beschouwing worden ingegaan. Hier willen wij besluiten met de opmerking, dat het verzamelen van gegevens voor het samenstellen van zo nauwkeurig mogelijke kansverwachtingslijnen voor een zo groot mogelijke verscheidenheid van predictoren en situaties één der belangrijkste taken is voor het wetenschappelijk werk op selectiegebied in de naaste toekomst.

HOOFDSTUK III

EFFECTIVITEIT EN RENDEMENT

Inleiding

In het vorige hoofdstuk is de gedachte ontwikkeld, dat de relatie tussen predictor en geschiktheidsbeoordeling kan worden weergegeven door een kansverwachtingskromme, waarbij bij iedere waarde van de predictor — of van een combinatie van predictoren — een bepaalde kans van slagen behoort. De vraag is nu, welke consequenties hieruit volgen voor de selectie. Daartoe zal eerst een concreet voorbeeld wat nader uitgewerkt worden. Dit biedt tevens gelegenheid, de relatie tussen predictor en geschiktheidsbeoordeling nog op een andere wijze te bezien, namelijk als een correlatieprobleem.

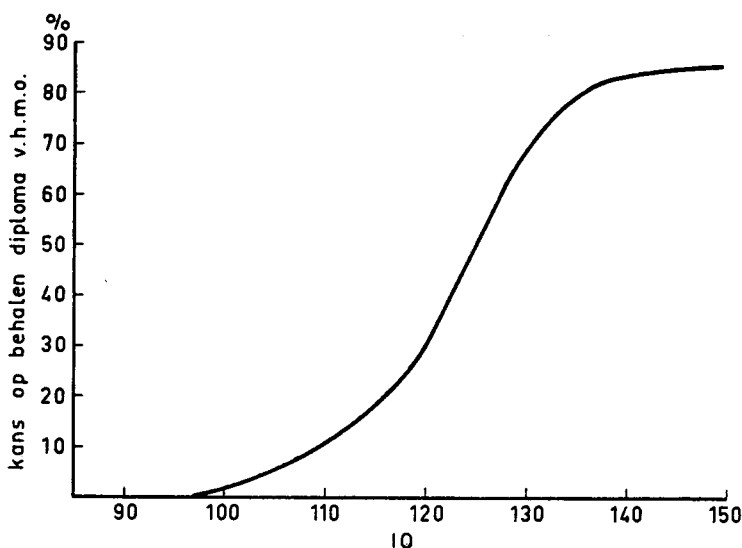
Geschiktheid voor het v.h.m.o.

Het concrete voorbeeld betreft de voorspelling van de geschiktheid voor het volgen van v.h.m.o. (dwz. gymnasium, lyceum, h.b.s., maar geen m.m.s. of handelsdagschool), op grond van de prestaties in een intelligentietest. Als criterium voor geschiktheid kiezen we het behalen van een einddiploma, ongeacht de tijd, die daarvoor benodigd is en ongeacht de redenen, die eventueel tot een eerder afbreken van de opleiding hebben gevoerd. Een dergelijk criterium is zeker niet zonder bezwaren. Het heeft echter als belangrijk voordeel, dat het op zeer eenvoudige wijze te beoordelen is.

De groep, waarop de beoordeling plaats vindt, zijn de kinderen, die zich voor toelating tot het v.h.m.o. aanmelden — zij kunnen dus eventueel reeds bij het toelatingsexamen worden afgewezen en gelden dan uiteraard als ongeschikt. Afbeelding 6 geeft een indruk van de relatie tussen IQ — intelligentiequotient, bepaald vóór de toelating — en kans van slagen. Daarbij moet worden opgemerkt, dat deze relatie van school tot school enigszins kan verschillen, zodat een gemiddeld beeld wordt gegeven, dat bovendien nog berust op een beperkt aantal waarnemingen.

Uit de afbeelding blijkt, dat er ditmaal een aantal gevallen is, waar men met zekerheid of aan zekerheid grenzende waarschijnlijkheid een voorspelling kan doen: de kans van slagen voor kinderen met een IQ lager dan 95 is nagenoeg nul. Dit hangt uiteraard samen met het feit, dat we hier te maken hebben met de groep opleidingen, die in het voortgezet onderwijs vóór de hoogste niveau-eisen stelt.

Zelfs zeer goede intellectuele begaafdheid is echter anderzijds geen garantie voor welslagen: uit de categorieën met de hoogste IQ's mislukt nog ongeveer één-zesde deel, dwz. in ieder geval een fractie, die veel te groot is om uitslui-



Afbeelding 6

tend uit oorzaken als ziekte, overlijden, vertrek wegens emigratie of grote welstandsval en dergelijke te kunnen worden verklaard.

Cut-off-scores

We willen nu eenvoudigheidshalve veronderstellen, dat afgezien van het toelatingsexamen het IQ het enige gegeven is, waarover we beschikken, als we bij de toelating tot het v.h.m.o. een zekere selectie zouden willen uitvoeren. De vraag is dan, waar we het beste de grens tussen geschikt en ongeschikt kunnen trekken. We kiezen dan een zogenaamde 'cut-off-score'. De keuze van deze score blijkt afhankelijk van de grondgedachten die we aan de selectie ten grondslag leggen.

We zouden bijvoorbeeld als richtsnoer kunnen kiezen, de wens om een zo klein mogelijk aantal onjuiste oordelen te geven. In dat geval ligt het voor de hand, de grenslijn te trekken op het punt, waar de kans van slagen 50% is, in concreto wordt dan dus de cut-off-score 125. In werkelijkheid is de situatie ingewikkelder dan hier wordt voorgesteld, men moet ook rekening houden met het feit, dat er andere opleidingsmogelijkheden zijn en met het feit, dat verschillende mogelijkheden vaak een ongelijke maatschappelijke waarde hebben. Op deze overwegingen zal later worden teruggekomen. Voorlopig bepalen we ons tot de gedachte, dat IQ 125 een cut-off-score is, die op grond van overwegingen van selectie-efficiency te verdedigen is.

Er zijn echter ook andere mogelijkheden: het is denkbaar, dat men minder geïnteresseerd is in het maken van een minimaal aantal fouten dan in een

combinatie van het verlangen naar een zekere selectie met de wens, om zo weinig mogelijk geschikten een kans te onthouden, zelfs als men daarvoor een offer in effectiviteit van selectie moet brengen. Denkbaar zou ook zijn een tegengestelde overweging, namelijk het verlangen om primair de ongeschikten te weren, zelfs als dit ten koste van de toelating van een aantal geschikten zou gaan. In de praktijk van onderwijsselectie zal men dit laatste verlangen niet zo spoedig tegenkomen, althans niet als voldoende plaatsen beschikbaar zijn, maar bij personeelselectie is dit uiteraard een bekend standpunt. Om nu te onderzoeken op welke wijze deze verschillende gezichtspunten invloed hebben op de selectie, moeten we nog geïnformeerd zijn over de relatieve frequentie, waarmee de verschillende IQ's voorkomen. Bekend is, dat de intelligentieverdeling van de totale bevolking bij benadering kan worden weergegeven door de beroemde klokvormige verdelingskromme van Gauss. Met deze verdeling als uitgangspunt komen we bij een cut-off-score van 125 tot de volgende berekening van de effectiviteit van het IQ als predictor.

IQ-gebied	P + (%)	f (%)	Onjuiste oordelen (%)		Juiste oordelen (%)	
— 95	0	37	0.00		37.00	
95+—100	1	13	.13		12.87	
100+—105	3	13	.39		12.61	
105+—110	7	11.7	.82		10.88	
110+—115	14	9.4	1.32		8.08	
115+—120	24	6.8	1.63		5.17	
120+—125	40	4.3	1.72		2.58	
125+—130	60	2.5		1.00		1.50
130+—135	75	1.3		.33		.97
135+—140	82	.6		.11		.49
140+—	84	.4		.06		.34
			6.01	+ 1.50	89.19	+ 3.30

P + (%): percentage geschikten in het betreffende IQ-gebied

f (%) percentage van de totale bevolking, dat in het betreffende IQ-gebied valt

De linker kolom 'onjuiste oordelen' geeft de percentages geschikten, die bij een cut-off score van 125 ten onrechte worden afgewezen, de rechter kolom geeft de percentages ongeschikten, die ten onrechte worden toegelaten. In de rubriek 'juiste oordelen' geeft de linker kolom de percentages der ongeschikten, die terecht worden afgewezen, de rechterkolom de percentages geschikten, die terecht worden toegelaten.

Men ziet in de eerste plaats, dat de predictor voor de totale bevolking een bijzonder hoog percentage juiste oordelen geeft, namelijk 92,5%. In zoverre zou men kunnen stellen, dat het IQ een uitzonderlijk goede predictor is. Anders gezegd: wanneer men een theorie van de geschiktheid voor v.h.m.o. wil opbouwen, die voor de bevolking als geheel geldig is, zal men zeer goed

gebruik kunnen maken van een model, dat nagenoeg geheel op het IQ gebaseerd is.

We zien ook: dat, uitgaande van IQ 125 als cut-off-score, ongeveer 5% van de totale bevolking tot het v.h.m.o. toegelaten zou kunnen worden. Van de toegelatenen zou dan $3.30/4.80 = 68\%$ geschikt zijn, in de zin, dat zij volgens de huidige normen een einddiploma zouden kunnen behalen. (In het tegenwoordige v.h.m.o. behaalt — rekening houdende met het feit, dat een aantal kandidaten reeds bij het toelatingsexamen wordt afgewezen — ongeveer 40% van de toelatingscandidaten t.z.t. een einddiploma.)

Als predictor, niet voor de schoolgeschiktheid van de totale bevolking, maar van het schoolsucces van de toegelatenen, is het IQ dus reeds aanzienlijk minder effectief dan bij de eerstgenoemde toepassing. De situatie wordt echter nog bedenkelijker, wanneer we ook letten op de geschikten, die bij de genoemde cut-off-score worden afgewezen: hun aantal is bijna tweemaal zo groot als dat dergenen, die terecht wél worden toegelaten. Op grond van deze overwegingen zal iemand die bij de selectie vooral beoogt zoveel mogelijk geschikten een kans te geven, bezwaren hebben tegen deze cut-off-score.

Uit de gerapporteerde gegevens kan men nu afleiden, welke consequenties de keuze van een andere cut-off-score heeft. Legt men deze bij IQ 120, dan blijven alle gegevens uit de tabel dezelfde, behalve die in de rij $IQ\ 120^+ - 125$, waar nu het percentage ten onrechte toegelatenen 2.58 en het percentage terecht toegelatenen 1.72 wordt. Tot het v.h.m.o. toegelaten wordt volgens dit criterium ca. 9% van de bevolking, namelijk 4,08% ongeschikten en 5,02% geschikten. Dwz. binnen de school is ca. 55% van de leerlingen geschikt, terwijl van alle geschikten nu ruim de helft tot de school wordt toegelaten. — Bij deze redeneringen wordt telkens stilzwijgend verondersteld, dat ieder, die volgens een bepaalde norm toegelaten kan worden, ook inderdaad het onderwijs gaat volgen! — Gaat men met de cut-off-score nog verder naar beneden, bijvoorbeeld naar IQ 115, dan ziet men, dat onder de toegelatenen de ongeschikten nu de meerderheid gaan vormen: de selectie is niet beter dan momenteel reeds het geval is.

Men kan ook onderzoeken, wat het effect is van strenge selectie, door bijvoorbeeld IQ 130 als cut-off-score te kiezen. Het percentage ongeschikten, dat dan nog door de zeef der selectie slipt, is klein: .5%, terwijl 2,3% geschikten wordt geaccepteerd. Dwz. in deze selecte groep zou ongeveer 82% der leerlingen volgens de huidige normen het diploma behalen, hetgeen in het licht van een streven naar het weren van ongeschikten een bevredigend resultaat zou kunnen worden genoemd. Thans wordt echter ongeveer drie-vierde deel van alle geschikten geweerd. Men ziet aan deze cijfers zeer duidelijk, hoe de beide doelstellingen: wering van ongeschikten en zoveel mogelijk geschikten een kans geven, elkaar in aanzienlijke mate uitsluiten, zelfs bij een predictor die ten aanzien van de bevolking als geheel zeer effectief zou kunnen worden genoemd.

Nu is in feite het betrekken van de kansverwachtingsgegevens op de bevolking als geheel weinig realistisch. Dit zou wél zinvol zijn bijvoorbeeld voor de situatie in Engeland, waar praktisch alle 11-jarige leerlingen van het lager onderwijs deelnemen aan een selectie-onderzoek voor de toelating tot de 'grammar schools'. In Nederland echter is dit geen gebruik. Degenen, die zich aanmelden voor toelating tot het v.h.m.o. vormen een reeds sterk voorgeselecteerde groep van de lagere school-bevolking. Kinderen met IQ lager dan 100 komen onder de toelatingscandidaten slechts in kleine aantallen voor.

We willen daarom de analyse nog eens herhalen, maar nu met als uitgangspunt de verdeling van de toelatingscandidaten naar IQ, zoals deze werd gevonden op een onderzoek in het kader van het toelatingsexamen van een groot lyceum. In eerste instantie wordt de cut-off score weer op 125 gesteld.

IQ-gebied	P + (%)	f (%)	Onjuiste oordelen (%)		Juiste oordelen (%)	
—105	3	5	.15		4.85	
105+—110	7	9	.63		8.37	
110+—115	14	13	1.82		11.18	
115+—120	24	16	3.84		12.16	
120+—125	40	20	8.00		12.00	
125+—130	60	17		6.80		10.20
130+—135	75	13		3.25		9.75
135+—	83	7		1.19		5.81
			14.44	+ 11.24	48.56	+ 25.76

In totaal wordt nu dus ongeveer drie-vierde deel van alle toelatingscandidaten juist beoordeeld. Zouden de leerlingen niet bewust op IQ worden geselecteerd, dan zal 40% het diploma behalen — de onderzochte groep is dus naar de landelijke maatstaf 'normaal'. Gaat men wel op IQ selecteren, dan zal bij een cut-off-score van 125 ruim één-derde (37%) worden toegelaten en van deze groep zal naar de huidige maatstaf 69% een einddiploma behalen. Dit resultaat wordt echter bereikt, door ongeveer 35% van alle geschikte kandidaten de toegang te weigeren.

Brengt men de cut-off-score op 120, dan blijkt 60% van de toegelatenen een einddiploma te behalen, terwijl nu nog 15% van de geschikten wordt buitengesloten. Bij een grenswaarde van 115 elimineert men nog ongeveer 4% van de geschikten, terwijl het percentage geslaagden voor het eindexamen nu 52 bedraagt. Verlaging van de cut-off-score tot 110 tenslotte maakt, dat het percentage geslaagden daalt tot 46, terwijl nog 2% van de geschikten niet tot het onderwijs wordt toegelaten.

In al de voorgaande gevallen is verondersteld, dat de normale beoordelings-procedure door de school, met inbegrip van het toelatingsexamen, onveran-

derd blijft bestaan en dat daarbij ook onveranderd dezelfde normen worden gehanteerd — in hoeverre deze laatste veronderstelling realistisch is, zal nog nader moeten worden gezien —. Tenslotte zij opgemerkt, dat het behalen van het einddiploma hier als het uitsluitende criterium voor schoolgeschiktheid is gehanteerd. In de werkelijkheid zal men vaak ook nog wel andere criteria willen hanteren, maar dit is meer voor de praktijk dan voor de theorie van de selectie van belang.

Beziet men nu de bovenstaande cijfers met het oog op de keuze van een realistische cut-off-score, dan komen we voor een opgave te staan, waarbij ongetwijfeld subjectieve inzichten een rol spelen. Als regel zal men het echter wel eens zijn met de stelling, dat de score die uit een oogpunt van doeltreffendheid van de beoordeling de beste is (IQ 125), voor de praktijk niet aanvaardbaar is, omdat veel te veel geschikten worden buitengesloten. Ook ten aanzien van IQ 120 zullen bij velen nog bedenkingen rijzen. Van de andere zijde benaderd zien we, dat het verschil tussen de grenzen IQ 115 en IQ 110 voornamelijk ligt in een verlies aan efficiency binnen de groep der geaccepteerden, zonder dat hiermee een noemenswaardige winst aan geschikten verkregen wordt. Men zou dus IQ 115 als een reële minimum-grens voor de selectie kunnen beschouwen: er worden dan heel weinig geschikten buitengesloten, terwijl er toch nog een merkbare winst aan doeltreffendheid is, vergeleken met het geval van geen selectie.

Correlatie

Het verband tussen predictor en beoordeling achteraf wordt dikwijls niet in de vorm van een kansverwachtingskromme gegeven, maar met behulp van een enkel getal, de correlatiecoëfficiënt. Ook de correlatiecoëfficiënt is een indicator, die ons een indruk geeft, binnen welke grenzen een bepaalde grootheid (het criterium) te voorspellen is, wanneer we een andere grootheid (de predictor) kennen.

Zouden we in een hypothetisch geval bijvoorbeeld vinden, dat voor een bepaalde functie de bedrijfsbeoordeling — op een schaal van nul tot tien — steeds met 0.1 punt stijgt met ieder punt dat de beoordeelde op een predictor meer behaalt en zouden we voorts weten, dat bij een predictorscore van 100 de bedrijfsbeoordeling 5,5 is, dan weten we voldoende om voor iedere predictorscore de bijbehorende bedrijfsbeoordeling te kunnen berekenen. In een dergelijk geval spreken we van een volmaakte lineaire afhankelijkheid, die wordt uitgedrukt door een correlatiecoëfficiënt van 1.00. Zou daarentegen in een ander geval ieder verband tussen predictor en criterium ontbreken, zodat bij elke waarde van de predictor dezelfde gemiddelde bedrijfsbeoordeling kan worden verwacht, dan drukken we dit uit in een correlatiecoëfficiënt van .00.

Als regel hebben we echter noch met het ene, noch met het andere geval te maken: er is wel enig verband, maar niet zodanig, dat een volledige voorspelling mogelijk is. Dit wordt dan aangeduid met een coëfficiënt, waarvan

de waarde tussen .00 en 1.00 ligt en wel des te dichter bij 1.00, naarmate de voorspelling met een geringere onnauwkeurigheid kan geschieden.

We maken voorts nog verschil tussen positieve en negatieve correlatie. In het geval van positieve correlatie behoort bij een grotere waarde van de ene grootte ook een grotere waarde van de andere grootte, bij negatieve correlatie is het omgekeerde het geval. Een voorbeeld van negatieve correlatie vormde in afbeelding 3 het verband tussen IQ en de kans op het verlaten van de U.L.O.-school zonder diploma.

Het voordeel van het werken met correlatiecoëfficiënten is uiteraard, dat met een enkel getal kan worden volstaan, in plaats van met een afbeelding. Hier staat echter als nadeel tegenover, dat de interpretatie van correlatiecoëfficiënten moeilijker is dan die van grafische voorstellingen. Bovendien zijn correlatiecoëfficiënten, in tegenstelling tot kansverwachtingkrommen, afhankelijk van de karakteristiek van de onderzochte groep ten aanzien van de onafhankelijke variabele.

Figuur 6 geldt bijvoorbeeld in principe zowel voor de bevolking als totaal, als voor de geselecteerde groep, die zich aanmeldt voor toelating tot het v.h.m.o. De correlatie *) echter tussen IQ en kans van slagen is voor de bevolking als geheel = .76, terwijl deze voor de als tweede voorbeeld besproken groep kandidaten voor een lyceum = .57 is. Dit blijkt samen te hangen met de geringere spreiding, die de tweede groep naar intelligentie heeft: naarmate een groep ten aanzien van een kenmerk homogener wordt, verliest dit kenmerk meer van zijn effectiviteit als predictor. Dit is gemakkelijk in te zien, als men bijvoorbeeld bedenkt, dat in het extreme geval, als de groep ten aanzien van het kenmerk geen enkele differentiatie meer vertoont, ook iedere mogelijkheid om dit kenmerk als predictor te gebruiken verloren is gegaan. In dat geval zal de correlatie met de te voorspellen grootte = .00 worden.

Correlatie en effectiviteit van selectie

Zoals bij de kansverwachtingkrommen is geschied, kan ook bij de correlatiecoëfficiënten het verband met effectiviteit van de selectie nader worden onderzocht. Daarbij blijkt een drietal grootheden een essentiële rol te spelen.

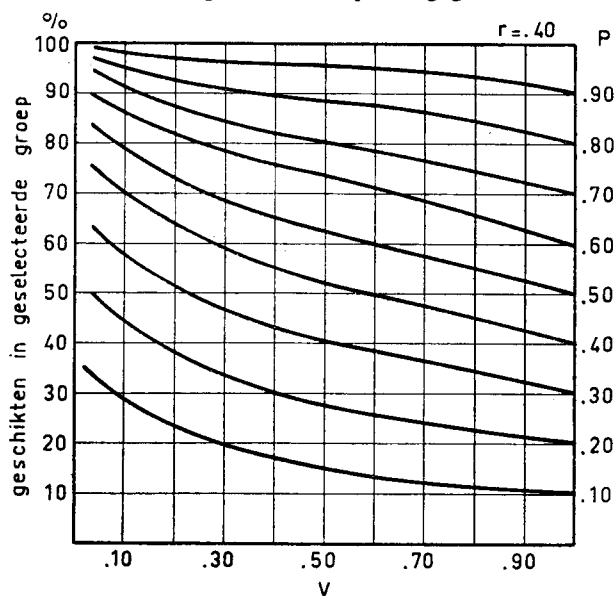
In de eerste plaats is dit uiteraard de grootte van de correlatie tussen de predictor en het criterium, uitgedrukt in een correlatiecoëfficiënt. In de tweede plaats een karakteristiek van het materiaal, waarop het onderzoek moet worden verricht. Hiervoor kan men gebruiken het percentage personen, dat geschikt is voor de functie, waarvoor wordt geselecteerd. Dit kenmerk wordt in de Amerikaanse literatuur wel de '*antecedent probability*' genoemd, wij zullen het als toevalskans betitelen. Men kan deze evalueren, wanneer men beschikt over gegevens betreffende geschiktheidsbeoordelingen, waarbij nog geen gebruik werd gemaakt van de predictor die wij willen bestuderen. Indien bijvoorbeeld voor een bepaalde functie tot dusver alleen selectie op grond van diploma's, referenties en de indruk uit een sollicitatiegesprek heeft plaats ge-

*) Deze is hier bepaald als een punt-biseriale correlatie.

vonden, hetgeen meebracht, dat in de praktijk 60% van de aangestelden bleek te voldoen en men overweegt nu om, met handhaving van de overige fasen uit de selectieprocedure, een psychologisch onderzoek te laten instellen, dan is voor de beoordeling van de effectiviteit van de predictoren die in dit onderzoek worden gebruikt de toevalskans = 60%.

De derde grootheid waarmee wij rekening moeten houden, heeft betrekking op de cut-off-score. Omdat wij nu niet meer een concreet geval bespreken, moet deze op algemene wijze worden aangegeven. Dit kan geschieden door middel van het percentage onderzochten, dat uiteindelijk voor aanstelling wordt aanbevolen. Uiteraard worden daarmee bedoeld degenen, die op de predictor de hoogste scores behalen. Dit kenmerk wordt wel aangeduid met de naam *selectieverhouding*.

Gewapend met gegevens over de toevalskans (P) en de selectieverhouding (V), kan men nu de betekenis van de correlatiecoëfficiënt (r) gaan interpreteren in termen van het percentage geschikten, dat in de groep positief geselecteerden mag worden verwacht. Voor een goed begrip moet er nog op worden gewezen, dat de relatie tussen de drie grootheden en het percentage geschikten ook weer een kansverwachting is. Dwz. als bijvoorbeeld wordt gevonden, dat bij een bepaalde combinatie van r, P en V het percentage geschikten 90 bedraagt, betekent dit niet, dat men steeds 90% geschikten in de geselecteerde groep zal aantreffen, maar dat als gemiddelde van de percentages in een aantal op deze wijze geselecteerde groepen de waarde 90% zal worden gevonden. Door twee Amerikaanse onderzoekers TAYLOR en RUSSELL, is het verband tussen r, P en V en het percentage geschikten in de geselecteerde groep voor een groot aantal gevallen berekend. Enkele van hun uitkomsten worden in de afbeeldingen 7 t/m 9 samengevat. Daarbij is uitgegaan van drie waarden voor

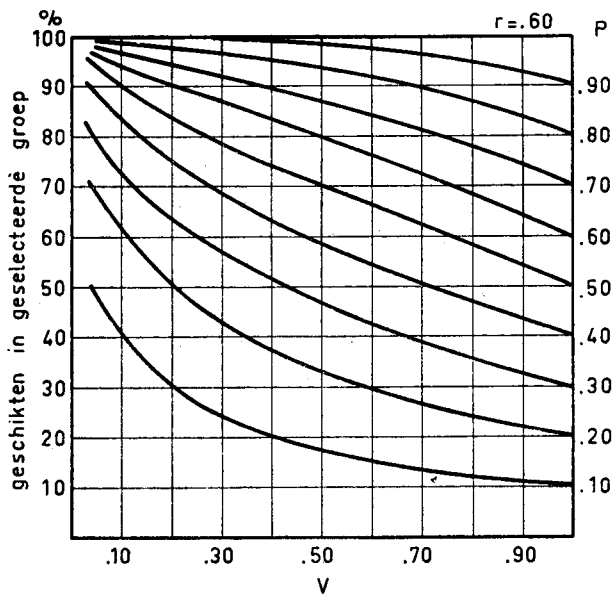


Afbeelding 7

de correlatie tussen predictor en criterium: .40, .60 en .80. Waarden hoger dan .80 zal men in de praktijk maar zeer zelden aantreffen. Als regel wordt een validiteit van .60 voor de totaliteit van het onderzoek reeds als gunstig beschouwd. Waarden lager dan .40 komen in de praktijk herhaaldelijk voor, zij hebben echter alleen betekenis voor praktische toepassing, wanneer men op zeer grote schaal selecteert terwijl bovendien de winst, verkregen door een kleine verbetering van de validiteit van de onderzoekprocedure, zeer veel groter is dan de vermeerdering van de kosten van onderzoek.

Bij bestudering van deze afbeeldingen valt vooral op, hoe belangrijk de rol van het aantal geschikten in het uitgangsmateriaal (de 'populatie') is.

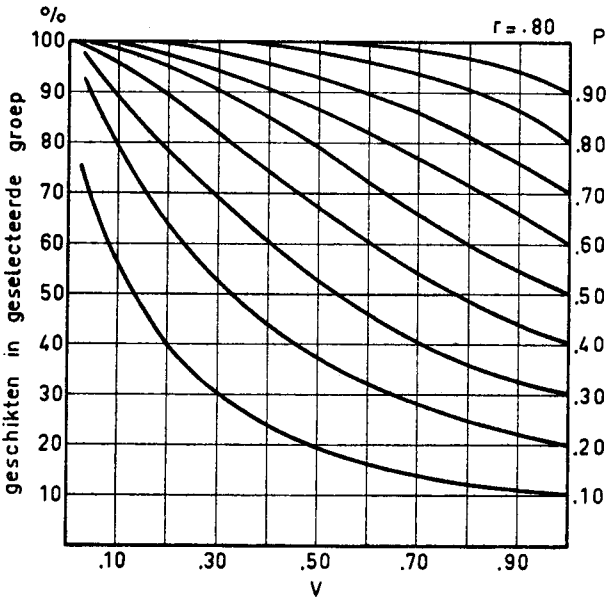
Bij een laag percentage geschikten in het aanbod, bijvoorbeeld 20% — een waarde, die in de praktijk zeker wel voorkomt — benadert men zelfs met een zeer valide selectieprocedure en met een zeer scherp gekozen selectieverhouding nog lang niet de 100% geschikten. Omgekeerd, als het percentage geschikten in het aanbod hoog is, is reeds de toevalskans op de aanstelling van geschikten zo groot, dat zelfs een zeer valide selectiemethode geen grote winst meer kan opleveren.



Afbeelding 8

Van het voorgaande zullen nu enkele toepassingen worden gezien. Stel het geval, dat een bedrijf uit een groot aanbod van kandidaten 10 plaatsen van één functie wil bezetten en dat men als bijkomende eis stelt, dat de statisti-

sche verwachting is, dat 9 van de 10 aanbevolen sollicitanten geschikt zullen zijn. De selectie vindt plaats met behulp van een methode, waarvan bekend is dat de correlatie van de totaalscore met beoordelingen van gebleken geschiktheid voor deze functie $r = .60$ is.



Afbeelding 9

De vraag is nu, welke selectieverhouding moet worden gekozen, om aan de bijkomende voorwaarde te voldoen. Deze verhouding is uiteraard afhankelijk van het percentage geschikten in het aanbod. De relevante gegevens kunnen worden ontleend aan afbeelding 8, zij worden in de hierna volgende tabel nog eens samengevat. Kent men de selectieverhouding, dan volgt uit het gegeven, dat men 10 personen wil aanstellen, hoeveel kandidaten onderzocht zouden moeten worden om de verlangde kwaliteit te bereiken. Dit aantal is eveneens in de tabel (rij A) aangegeven.

P	10	20	30	40	50	60	70	80	90%
V	.01	.015	.025	.045	.10	.20	.40	.70	1.00
A	1000	667	400	222	100	50	25	15	(10) kandidaten

Hieruit blijkt, dat bij een laag percentage geschikten in het aanbod de selectie op een zodanig massaal aanbod zou moeten worden uitgevoerd, dat hieraan in de praktijk niet te denken valt. In normale gevallen zal het aantal te onderzoeken personen in het algemeen niet meer dan 5 à 10 maal groter kunnen zijn dan het aantal aan te stellen personen. Vaak zal zelfs dit veelvoud niet te bereiken zijn. Dit betekent, dat men in gevallen van een relatief gering percentage geschikten in het aanbod dus voorzichtig moet zijn met het stellen

van bijkomende eisen aan de kwaliteit van de geselecteerde groep.

Rekening houdende met het gegeven, dat men hoogstens het 5 à 10-voud van het aantal aan te stellen personen zal onderzoeken, zodat dus de selectieverhouding maximaal .10 à .20 zal bedragen, kunnen we uit afbeelding 8 ook weer afleiden, welk percentage geschikten men in de geselecteerde groep mag verwachten, afhankelijk van het percentage geschikten in het aanbod:

P	10	20	30	40	50	60	70	80	90
V=.10	39	60	74	83	90	94	97	98	100% geschikten
V=.20	30	50	64	75	84	90	95	97	99% geschikten

Hieruit blijkt dus, dat het verlangen, ca. 90% geschikten in de geselecteerde groep te hebben, alleen te realiseren is, als het percentage geschikten in het aanbod minstens in de orde van grootte van 50% ligt, en dan nog alleen, wanneer men bereid is scherp te selecteren en dus veel personen te onderzoeken.

Men kan nu de vraag nog enigszins anders stellen. We nemen aan dat een bedrijf 10 geschikte kandidaten voor een functie zoekt en bereid is een groter aantal aan te stellen, zodanig dat de kansverwachting ten aanzien van het aantal geschikten 10 zal bedragen. Dit aantal aan te stellen kandidaten is uiteraard weer afhankelijk van de selectieverhouding en hieruit kunnen we dan tevens afleiden, hoeveel personen onderzocht zouden moeten worden. De hierna volgende tabel geeft hiervan een overzicht. Een open plaats in deze tabel betekent, dat evenveel personen aangesteld zouden moeten worden als aangegeven door de eerstvolgende waarde lager in de betreffende kolom, echter na een scherpere en dus onnodig kostbare selectie.

P		10	20	30	40	50	60	70	80	90
V = .10	Aanst.	26	17	14	12	11				
	Ondz.	260	170	140	120	110				
V = .20	Aanst.	33	20	16	13	12		10		
	Ondz.	165	100	80	65	60		50		
V = .30	Aanst.	40	23	17	14		11			
	Ondz.	133	77	57	47		36			
V = .40	Aanst.	48	26	19	16	13	12		10	
	Ondz.	120	65	48	40	33	30		25	
V = .50	Aanst.	56	30	21	17	14		11		
	Ondz.	112	60	42	34	28		22		
V = .60	Aanst.	65	34	23	18	15	13			
	Ondz.	108	57	38	30	25	22			
V = .70	Aanst.	74	38	26	20	16		12		10
	Ondz.	106	54	37	29	23		17		14
V = .80	Aanst.	83	42	28	21	17	14		11	
	Ondz.	104	53	35	26	21	18		14	
V = .90	Aanst.	91	45	30	23	19	15	13	12	
	Ondz.	101	50	33	26	21	17	14	13	
Geen selectie										
Aanstellen		100	50	33	25	20	17	14	13	11

De algemene regel die in bovenstaande tabel vervat is luidt, dat men naarmate men minder scherp selecteert, minder kandidaten behoeft te onderzoeken om toch een bepaald aantal geschikten te vinden, maar dat men in dat geval ook meer ongeschikten aanstelt. Is evenwel het percentage geschikten in het aanbod hoog, dan is de winst door het elimineren van ongeschikten gering. Bij $P = 70\%$ zal men zonder selectie 14 personen moeten aanstellen, om 10 geschikten te mogen verwachten. Wil men er vrijwel zeker van zijn, dat alle aangestelden ook geschikt zijn, dan zal men — bij een validiteit van de selectieprocedure van .60 — een selectieverhouding van .20 moeten aanhouden en dus 50 personen onderzoeken. Bij een nog hogere toevalskans wordt de door selectie te bereiken winst relatief nog geringer. Tenzij er zeer grote voordelen te verwachten zijn indien de enkele ongeschikten nog geëlimineerd worden, zal men dus, als de bestaande selectieprocedure reeds tot ongeveer 70% of meer geschikten voert, nauwelijks behoeven te overwegen om door het inschakelen van verdere predictoren nog tot een verbetering van de selectie te komen.

Is het percentage geschikten in het uitgangsmateriaal laag, dan is geen van de alternatieven, waarvoor men wordt geplaatst, erg aantrekkelijk.

Bij $P = .20$ zal men zonder verdere selectie 50 personen moeten aanstellen, om uiteindelijk 10 geschikten over te kunnen houden. Is men bereid 100 personen te laten onderzoeken, dan kunnen daaruit eventueel de 20 besten worden gekozen, waaronder men dan eveneens 10 geschikten mag verwachten, hetgeen na een groot opgezet onderzoek toch nog slechts een rendement van 50% binnen de geselecteerde groep betekent.

In dergelijke gevallen kan men soms echter toch nog tot een bevredigende werkwijze komen, indien men de selectie over een aantal ronden verdeelt, waarbij telkens slechts een eenvoudige, gemakkelijk toe te passen predictor, of serie predictoren wordt gebruikt. Iedere ronde draagt dan iets bij tot de concentratie van het aantal geschikten, tot men tenslotte een 'antecedent probability' bereikt, die voor de eindselectie voldoende is.

Een dergelijke werkwijze is welbekend als het gaat om de selectie van één functionaris, in het algemeen voor een hogere functie, maar zij kan ook bij selectie op grotere schaal toegepast worden.

Het rendement van de selectie

In het voorgaande is reeds een enkele maal aangeduid, dat selectie of althans voortgezette selectie onder bepaalde omstandigheden economisch nauwelijks te verdedigen is. Dit probleem van het economisch rendement van selectie kan nu wat nader worden gezien. Voorlopig stellen we het probleem eenvoudig: het onderzoek kost per onderzochte a gulden, terwijl de aanstelling van een ongeschikte b gulden meer kost dan de betrokkene door zijn prestaties positief inbrengt. In de praktijk kan de grootheid a gemakkelijk worden gekend, terwijl b in het algemeen slechts door schatting kan worden benaderd, omdat er bij ongeschikte functionarissen vaak sprake is van onzichtbare

schade. Naar de orde van grootte is echter ook b , zowel in het geval van ongeschikte werknemers in een bedrijf als in het geval van ongeschikte leerlingen in een opleiding, wel bekend, zodat de berekening in vele gevallen concreet kan worden gemaakt. Ter toelichting van de berekening van het rendement denken we een concreet geval, waarin de toevalskans $P = 40\%$. Nu kan men bijvoorbeeld, als men weer streeft naar 10 geschikte functionarissen:

	onderzoeken	aanstellen	waarvan ongeschikt
1. zonder selectie		25	15
2. $V = .50$	34	17	7
3. $V = .30$	47	14	4
4. $V = .20$	65	13	3

Voor de goede orde wijzen wij er nog op, dat bovenstaande getallen alleen gelden voor een predictor met een validiteit van .60. Voor andere validiteitswaarden kan men t.a.v. $r = .40$ en $r = .80$ de gegevens afleiden uit de afbeeldingen 7 en 9, terwijl tussenliggende waarden door interpolatie kunnen worden gevonden (in de oorspronkelijke publicatie van TAYLOR en RUSSELL kan men gegevens vinden voor waarden van r tussen .00 en 1.00, telkens met .05 opklimmend). Voorts dient er ook nog op te worden gewezen, dat de berekening veronderstelt, dat men kan putten uit een voldoende aanbod, om eventueel betrekkelijk scherp te kunnen selecteren.

Men ziet dan uit de gegevens, dat in het onderhavige geval selectie economisch loont in de volgende situaties:

2 loont t.o.v. 1 als $34 a < 8 b$	(In geval 2 zijn er 8 ongeschikten minder
3 loont t.o.v. 2 als $13 a < 3 b$	dan in geval 1 en 34 personen meer te
4 loont t.o.v. 3 als $18 a < b$	onderzoeken.)

Dus wanneer de aangerichte schade per ongeschikte ruim viermaal zo groot is als de kosten van extra onderzoek per kandidaat, loont selectie reeds, mits daarbij een predictor van redelijk goede validiteit ter beschikking staat. In dit concrete geval loont het dan zelfs aanstonds, een selectieverhouding van .30 te kiezen, dus 10 kandidaten te laten onderzoeken voor elke 3 aan te stellen functionarissen. Een nog scherpere selectieverhouding loont daarentegen eerst, als de te verwachten schade zeer veel groter is dan de kosten per onderzoek.

Voert men dezelfde berekening ook uit voor het geval van $P = .20$, dan blijkt selectie reeds te lonen als de schade per ongeschikte driemaal zo groot is als de kosten per onderzoek, terwijl een selectieverhouding van .20 nu reeds loont als de schade per ongeschikte $4\frac{1}{2}$ maal zo groot is als de kosten per onderzoek. Terwijl dus in het voorgaande opgemerkt kon worden, dat selectie in het geval van een lage 'antecedent probability' weinig bevredigend is, omdat men in de geselecteerde groep veel ongeschikten overhoudt, zelfs als men vrij scherp selecteert, blijkt nu, dat selectie juist in deze gevallen uit

economisch standpunt het eerste verantwoord is. We krijgen hier te maken met een merkwaardige anomalie, want in dergelijke gevallen is vaak de beoordelaar weinig bevredigd door het resultaat van de selectie, terwijl toch het economisch rendement groter is dan in gevallen, waarin door een gunstiger toevalskans het percentage geschikten in de geselecteerde groep groter is.

Gevallen, waarvoor de bovenstaande redenering geldig is, komen met name in de praktijk van personeelselectie frequent voor. Er zijn echter ook gevallen, waarin met meer factoren rekening moet worden gehouden. We hebben gezien, dat selectie fouten naar twee zijden maakt: er worden ongeschikten geaccepteerd en geschikten afgewezen. Hoe geringer we de fout van de ene soort maken, des te groter wordt de fout van de andere soort. Ook daardoor kan schade ontstaan. In het kader van onderwijsselectie zal bijvoorbeeld het feit, dat iemand ten onrechte niet tot een bepaalde opleiding wordt toegelaten, kunnen inhouden, dat de betrokkene bepaalde potenties niet zal weten te realiseren. Nog afgezien van de nadelen die daaruit voor hem persoonlijk voortvloeien, kan dit een maatschappelijke schadepost betekenen. Gemiddeld genomen zal iemand, die bijvoorbeeld v.h.m.o. heeft voltooid, in de loop van zijn leven wel een grotere bijdrage tot het nationaal inkomen leveren dan de gemiddelde bezitter van een m.u.l.o.-diploma. Ook in andere gevallen kan het afwijzen van geschikten voor de instantie die opdracht tot onderzoek bij selectie geeft, een schadepost betekenen. De grootte hiervan stellen we symbolisch op c gulden.

Wanneer we thans nog eens het voorbeeld bezien, dat in eerste instantie ter toelichting van de berekening heeft gediend, dan zien we, dat bij een selectieverhouding van .50 een groep van 34 personen moest worden onderzocht. Bij $P = 40\%$ bevinden zich in deze groep ongeveer 14 geschikten, terwijl de geselecteerde groep 10 geschikten bevat. Er zijn dus 4 geschikten ten onrechte afgewezen. Op analoge wijze berekenen we dat bij selectieverhouding .30 het aantal afgewezen geschikten 11 bedraagt, bij selectieverhouding .20 is dit tot 16 gestegen.

De berekening wordt nu, dat

2 loont t.o.v. 1 als $34a + 4c < 8b$

3 loont t.o.v. 2 als $13a + 7c < 3b$

4 loont t.o.v. 3 als $18a + 5c < b$

De situatie wordt nu geheel anders dan in het eerste geval. Veelal zal a wel aanzienlijk kleiner zijn dan b . Maar het is heel goed denkbaar, dat b en c in dezelfde orde van grootte liggen, dwz. dat de schade, aangericht door een onjuiste beslissing, naar beide kanten ongeveer even groot is. In dat geval zullen de procedures 3 en 4 verwerpelijk zijn, alleen de betrekkelijk milde selectie, met een selectieverhouding .50, is dan waarschijnlijk nog wel acceptabel.

Voor andere waarden van P kan men analoge redeneringen opzetten. Zij zullen hier niet worden weergegeven; met de benodigde gegevens, die aan de

afbeeldingen 7 t/m 9 ontleend kunnen worden, kan de lezer desgewenst deze gevallen zelf onderzoeken. Wij vermelden alleen als conclusie, dat naarmate P hoger wordt, de kans dat scherpe selectie loont kleiner wordt. Het spreekt voorts wel vanzelf, dat selectie eerder loont, naarmate de onderzoekprocedure meer valide is, omdat daardoor naar beide kanten het aantal fouten geringer wordt.

De beschouwingen over het rendement van selectie zijn daarmee niet ten einde. In de realiteit stuit men vaak op veel gecompliceerder situaties, waarin rekening moet worden gehouden met verschillende mogelijkheden, zodat sprake is van distribuerende selectie. De theorie die voor dergelijke gevallen ontwikkeld is, zal ons in het volgende hoofdstuk bezig houden.

HOOFDSTUK IV

OPERATIONS RESEARCH

Inleiding

De situaties, die we tot dusver hebben gezien, behoren tot de categorie van de enkelvoudige selectie. De beoordeling geldt één functie, waarvoor men al dan niet geschikt wordt geacht. De selectiepraktijk kent echter ook andere situaties. Er zijn dan verschillende functies in het geding, en de vraag is niet in de eerste plaats, of er in absolute zin sprake is van geschiktheid, maar voor welke functie een gegeven persoon primair in aanmerking komt, respectievelijk, welke functie hij het beste kan kiezen. In een geval van schoolkeuze is de vraagstelling niet: v.h.m.o. of niets, maar het maken van een rangorde van verschillende mogelijkheden: v.h.m.o., u.l.o. en eventueel nog andere onderwijsvormen. In het kader van personeelsselectie correspondeert hiermee het geval, waarin men een sollicitant graag zal willen aanstellen, maar er nog een keuze moet worden gemaakt tussen verschillende functies waarin plaatsen beschikbaar zijn. In dergelijke gevallen van differentiatie en van distribuerende selectie doen zich problemen voor, die bij enkelvoudige selectie geen rol spelen.

Keuze tussen onderwijsmogelijkheden

Ter toelichting van de problematiek gebruiken we weer een concreet geval, waarbij op grond van een predictor, in casu het IQ, een aanbeveling moet worden gedaan inzake de keuze tussen v.h.m.o. en u.l.o. Dit is dus nog een betrekkelijk eenvoudig geval, we veronderstellen, dat het v.h.m.o. gerepresenteerd wordt door één schooltype (lyceum), terwijl de u.l.o.-school behoort tot de categorie, waarover in het tweede hoofdstuk een grafiek van de kansverwachting werd weergegeven. Andere keuzemogelijkheden (bijvoorbeeld lager nijverheidsonderwijs) blijven buiten beschouwing.

Een eerste probleem is nu, dat met name de grafiek van de kansverwachting ten aanzien van u.l.o. misleidend is. De gegevens hebben namelijk betrekking op leerlingen, die bij een gegeven IQ u.l.o. hebben gekozen. De gebruikelijke praktijk van ons onderwijs is echter die van een trapsgewijze selectie, waarbij de leerlingen met de beste schoolresultaten zoveel mogelijk in de richting van v.h.m.o. worden gestuurd, aan de tweede laag u.l.o. wordt aanbevolen, terwijl de overblijvenden naar lager nijverheidsonderwijs en v.g.l.o. worden gedirigeerd. Natuurlijk is deze voorstelling van zaken schematisch: er wordt met name op de tweede en derde trap ook wel rekening gehouden met de belangstelling van de leerlingen en met de fijnere structuur van zijn capaciteiten, maar de zeefwerking is toch een reëel feit.

Nu is onze predictor, het IQ, als regel niet het instrument, dat voor de zeefwerking van het onderwijs wordt gehanteerd. Deze geschiedt eerder op basis van prestaties in het lager onderwijs. Hierin zijn, naast de intelligentie, ook factoren als werkhouding, belangstelling, geheugen, steun van het milieu en

dergelijke verdisconteerd. We vinden dus bij gelijk IQ leerlingen, die naar het v.h.m.o. en andere, die naar het u.l.o. zullen gaan. Zouden we beide groepen beoordelen op hun kans van slagen, indien zij u.l.o. zouden kiezen, dan is de kansverwachting bij de eerstgenoemde groep allicht gunstiger dan bij de tweede. Met andere woorden, de kansverwachtingsschijf, die voor de relatie IQ-behalen m.u.l.o.-diploma werd gevonden, is beïnvloed door het feit, dat het v.h.m.o. een deel van de meest geschikte leerlingen uit iedere IQ-groep wegzuigt. Willen we bij een gegeven individu, dat nog niet heeft gekozen, a priori een uitspraak doen over de kans van slagen in het u.l.o., dan zullen we voor dit wegzuig-effect een correctie moeten aanbrengen.

Strikt genomen moet voorts, zowel ten aanzien van het v.h.m.o. als van u.l.o., nog een correctie worden aangebracht voor het feit, dat zich in de groep leerlingen, die voor eenvoudiger onderwijsvormen opteren, ook kinderen bevinden, die in staat zouden zijn een v.h.m.o., c.q. mulo-diploma te behalen, maar dit effect is kwantitatief minder belangrijk. Het zal hier buiten beschouwing blijven.

We wijzen er nog op, dat in het kader van personeelsselectie geheel analoge wegzuigeffecten kunnen optreden, die dan eveneens aanleiding kunnen geven de kansverwachting vóór de beslissing te corrigeren.

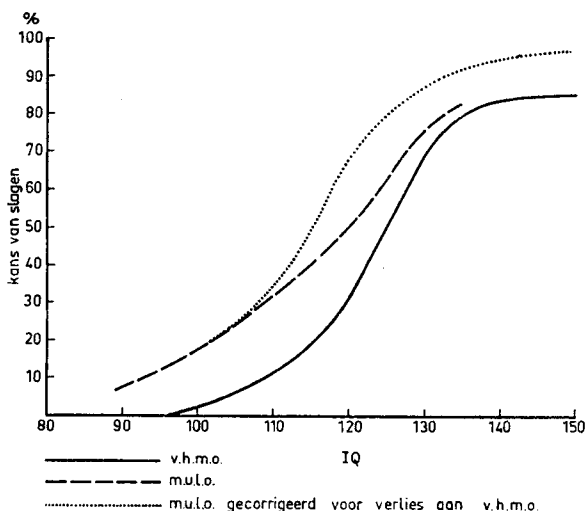
Voor een nauwkeurige correctie zou men over zeer gedetailleerde gegevens moeten beschikken over de schoolcarrière van v.h.m.o.-leerlingen met een gegeven IQ. Zolang deze niet beschikbaar zijn, moet met een globale benadering worden volstaan. In afbeelding 10 vindt men een aantal gegevens verenigd: a) de kansverwachting bij keuze van v.h.m.o. (dezelfde curve als in figuur 6), b) de kansverwachting bij keuze van u.l.o. (de gestippelde lijn uit figuur 3), c) een gecorrigeerde *) kansverwachting voor u.l.o. De lijnen a) en c) zullen als uitgangspunt voor de thans volgende analyse worden gebruikt.

Uiteraard zullen we voor het doen van een aanbeveling inzake de keuze weer moeten beschikken over een criterium. We gaan er van uit, dat dit van economische aard is. Voorts veronderstellen we, dat we daarbij uitsluitend zullen letten op het economisch voordeel van de persoon, die een keuze moet maken. **)

*) De correctie is gebaseerd op de volgende veronderstellingen:

1. Op grond van gegevens uit enkele schoolonderzoekingen wordt geschat, dat uit de groep IQ 105–110 het aantal u.l.o.-candidaten $10 \times$ zo groot is als het aantal v.h.m.o.-candidaten. In de groep 110+ – 115 is de verhouding 5 : 1, in de groep 115+ – 120 2 : 1, in de groep 120+ – 125 1 : 1, bij hogere IQ's kan de proportie van u.l.o.-leerlingen verwaarloosd worden.
2. Binnen elk van deze IQ-groepen is de kans op het behalen van een u.l.o.-diploma bij de v.h.m.o.-candidaten (apriori) tweemaal zo groot als bij de u.l.o.-candidaten.
3. De grote meerderheid van de leerlingen met hoge IQ's zou in staat zijn een u.l.o.-diploma te behalen.

**) En passant kan er op worden gewezen, dat dit een illustratie is van het *dualiteitsprincipe van de selectie* (vgl. het eerste hoofdstuk). Hetzelfde probleem kan worden behandeld als een keuze tussen personen of tussen situaties.



Afbeelding 10

Als maatstaf voor de beoordeling van het economisch voordeel ligt voor de hand, het inkomen, dat iemand direct of indirect (bijv. in de vorm van pensioen) uit arbeid mag verwachten, wanneer gegeven is, dat hij een bepaalde vooropleiding heeft. Uiteraard is deze verwachting niet meer dan een constatering over een statistisch gemiddelde, waaromheen een zeer aanzienlijke spreiding optreedt. Bovendien moeten we ons in feite baseren op een situatie in het verleden: het is natuurlijk lang niet uitgesloten, dat, door een wijziging in de maatschappelijke waardering van bepaalde vooropleidingen of door een wijziging in de beloningsverhoudingen tussen de beroepen, het gemiddeld te verwachten inkomen in de toekomst anders zal zijn dan in het verleden. Zolang er echter geen mogelijkheid bestaat, deze trend te voorspellen, zal men zich bij het nemen van een beslissing wel noodgedwongen op de actuele beloningsverhoudingen moeten baseren.

Om geen complicaties te krijgen met problemen van geldontwaarding en dergelijke, gebruiken we geen absolute maatstaf, maar we kiezen als eenheid het gemiddeld te verwachten voordeel van één der in geding zijnde mogelijkheden — in casu het gemiddeld te verwachten inkomen van bezitters van een v.h.m.o.-diploma.

Het voorbeeld zal thans met geschatte cijfers moeten worden voltooid. In werkelijkheid beschikken we namelijk nog niet over een goed inzicht in de economische waarde van verschillende vooropleidingen, hoewel misschien de gegevens die bij de volkstelling 1960 zijn verzameld een voldoende basis zouden kunnen geven voor althans een eerste schatting dienaangaande.

Wanneer in het geding is de keuze tussen v.h.m.o. en u.l.o., krijgen we te maken met vier economische waarden: die van het bezit van diploma v.h.m.o., die van het gevolgd hebben van v.h.m.o. zonder het eindexamen te behalen, die van het bezit van het m.u.l.o.-diploma en tenslotte die van het volgen van u.l.o., zonder het diploma te behalen. Deze waarden duiden we achtereenvolgens aan met a_1 , b_1 , a_2 , b_2 , waarbij $a_1 = 1$, omdat deze waarde als eenheidsmaatstaf wordt gekozen. Als de kans van slagen op v.h.m.o. bij een bepaalde waarde van de predictor x bedraagt (in dit geval wordt x niet als een percentage gegeven, maar als de daarmee overeenkomende fractie tussen 0 en 1), dan is de economische waarde van de keuze van v.h.m.o. gelijk aan $y_1 = a_1 x + b_1 (1 - x) = (1 - b_1) x + b_1$.

De vergelijking voor de economische waarde van de keuze van u.l.o. luidt analoog $y_2 = a_2 x + b_2 (1 - x) = (a_2 - b_2) x + b_2$. Noemen we de verhouding $b_2/a_2 = p_2$, dan kunnen we de vergelijking voor de economische waarde van de keuze van u.l.o. ook schrijven als

$$y_2 = a_2 \{ (1 - p_2) x + p_2 \}$$

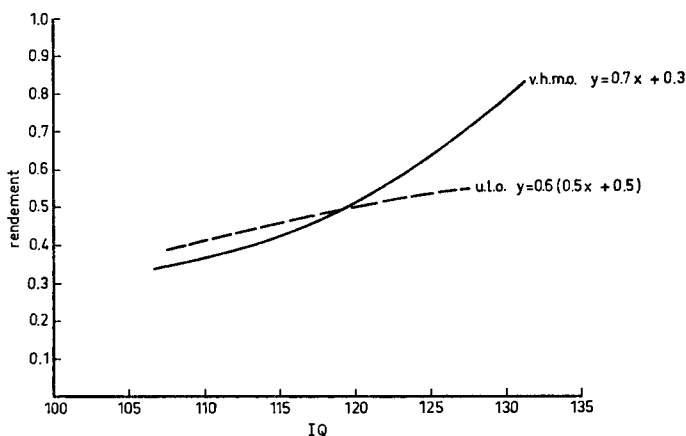
waardoor de analogie met de vergelijking voor y_1 duidelijker tot uitdrukking komt.

In het licht van de huidige maatschappelijke verhoudingen is wel duidelijk, dat de verwachting ten aanzien van het gemiddelde inkomen — over het gehele arbeidsleven — voor de bezitter van een v.h.m.o.-diploma hoger is, dan voor de bezitter van een m.u.l.o.-diploma. Schattenderwijs stellen we de verhouding op 5 : 3, zodat $a_2 = .6$. Voorts mogen we wel aannemen, dat bij het bezoeken van een bepaalde school het voordeel van het behalen van het diploma op het v.h.m.o. groter is dan op het u.l.o. Deze veronderstelling is vooral gebaseerd op de overweging, dat het bezit van een v.h.m.o.-diploma de weg opent naar de academische studies. We kiezen voor v.h.m.o. de verhouding tussen wel en niet behalen van het diploma als 10 : 3 ($b_1 = .3$), terwijl we voor u.l.o. de verhouding 2 : 1 ($p_2 = .5$) nemen.*).

De waarde voor x bij een gegeven predictorwaarde kunnen we ontlelen aan afbeelding 10. Als we voor een aantal punten uit afbeelding 10 de bijbehorende waarden van y_1 en y_2 berekenen, kunnen we deze in een grafische voorstelling uitzetten en vervolgens, door deze punten onderling te verbinden, de lijnen tekenen, die voor iedere waarde van de predictor de economische waarde van de keuze van v.h.m.o., respectievelijk u.l.o. aangeven.

Indien onze veronderstellingen de werkelijkheid zouden weergeven, dan blijkt uit afbeelding 11, dat beneden IQ 118 de keuze van u.l.o. uit een oogpunt van verwacht economisch nut de voorkeur verdient, terwijl boven IQ 119 de keuze van v.h.m.o. gunstiger perspectieven opent.

*) Deze verhoudingsgetallen zijn extremer dan door de realiteit gerechtvaardigd schijnt. Zij zijn echter zo gekozen, ter compensatie voor het feit, dat nog geen rekening is gehouden met een hierna te bespreken complicatie.



Afbeelding 11

In feite is het hier gebruikte model nog geen goede weerspiegeling van de werkelijkheid. Tot dusver hebben we steeds de fictie gehanteerd, dat ongeacht de waarde van predictor allen die geschikt zijn dezelfde economische waarde hebben. Zo ook degenen, die ongeschikt blijken. Deze veronderstelling is echter niet juist. In het geval van keuze van v.h.m.o. is wel bekend, dat iemand met een relatief laag IQ, die het diploma behaalt, gemiddeld genomen minder gunstige perspectieven heeft dan iemand met een hoger IQ. Bij een laag IQ is er meer kans op doubleren, meer kans, dat men de relatief gemakkelijkste richting (h.b.s.-A) kiest, meer kans, dat men daarna niet meer begint aan een relatief zware studie en daardoor ook een aanzienlijke kans, dat men onder de bezitters van een v.h.m.o.-diploma naar inkomen tot de achterhoede zal behoren. Uiteraard is ook dit alleen maar een uitspraak, die betrekking heeft op het gemiddelde geval, de individuele gevallen vertonen een grote mate van verspreiding rondom dit gemiddelde.

Als we met deze finesse ook kwantitatief rekening willen houden, zullen we moeten beginnen met een model te construeren, waardoor de relatie tussen predictorscore en rendement bij slagen op mathematische wijze wordt bepaald. Een eenvoudig model is dat van de lineaire relatie, waarbij het economisch nut recht evenredig met de predictorscore (hier het IQ) toeneemt. Of dit model de werkelijkheid voldoende benadert weten wij nog niet, dit kan alleen getoetst worden, als wij eerst over voldoende gegevens betreffende concrete gevallen beschikken. Als bijvoorbeeld zou blijken, dat bij hoge predictorscores (hoog intelligentiequotient) het economisch voordeel bij het behalen van het diploma niet meer toeneemt, maar integendeel misschien eerder zelfs enigszins afneemt, kan men niet volstaan met een lineair model — men zal dan moeten onderzoeken of wellicht een kwadratische relatie de werkelijkheid beter benadert.

Bij gebrek aan voldoende basisgegevens zal de gedachtengang thans niet verder concreet worden uitgewerkt. Beschikt men wel over deze gegevens, dan

zal men uiteindelijk ook weer een rendementscurve, analoog aan die van afbeelding 11, verkrijgen.

Operations research

Deze ontwikkeling van de theorie van de selectieproblemen, waarbij gebruik wordt gemaakt van mathematische modellen, voert ons op het gebied van de 'operations research' (besliskunde), dwz. de wetenschap, die zich ten doel stelt, het nemen van beslissingen — in het bijzonder van grote series beslissingen — te vereenvoudigen, door een kwantitatieve evaluatie van de consequenties van de verschillende alternatieven.

Ten aanzien van de selectieproblemen hebben vooral verschillende Amerikaanse onderzoekers, van wie hier CRONBACH en GLESER speciaal moeten worden genoemd, belangrijke bijdragen tot de theorie van de kwantitatieve benaderingswijze geleverd. Momenteel verkeert deze tak van wetenschap, ook in de Verenigde Staten, nog vrijwel geheel in het stadium van theoretisch onderzoek, omdat de benodigde externe gegevens, in het bijzonder die van het economisch rendement, nog maar mondjesmaat beschikbaar zijn.*) Voorlopig is de belangrijkste toepassing wel, dat het door deze theoretische ontwikkeling mogelijk is geworden allerlei hypothetische gevallen door te rekenen, zodat een dieper inzicht wordt verkregen in de invloed van verschillen in selectieratio of van verschillen in de kwantitatieve verdeling van de onderzochte personen over verschillende functies, in de invloed van de validiteit van de predictoren en van de kostprijs van het onderzoek etc. Vooral nu men met behulp van computers in korte tijd een groot aantal gevallen kan doorrekenen, heeft dit ook voor de praktijk belangrijke waarde. Bovendien ligt nu het theoretisch schema gereed, om — zodra de benodigde praktijkgegevens beschikbaar komen — een vérgaande kwantitatieve benadering van het selectieprobleem te hanteren.

Sequente selectie

In het kader van de recente theoretische ontwikkelingen heeft men ook veel aandacht geschonken aan de sequente selectie. Het principe hiervan is welbekend: het heeft voordelen, om een selectie trapsgewijze uit te voeren. Hierop berust uiteraard ook het gebruikelijke systeem van voorselecteren, bijvoorbeeld eerst op grond van formele kenmerken (leeftijd, opleiding, ervaring, e.d.), daarna eventueel op grond van een grafologische beoordeling en het vragen van referenties, om pas daarna op degenen, die deze voorselectie doorstaan hebben, de gehele, betrekkelijk kostbare verdere selectieprocedure toe te passen.

*) Na het schrijven van dit hoofdstuk verscheen de eerste Nederlandse studie waarin de theorie van 'Operations research' wordt toegepast op een concreet probleem van personeelsselectie: R. F. van Naerssen, *Selectie van chauffeurs*. Groningen 1963.

In de praktijk maakt men zich over de validiteit van een dergelijke werkwijze slechts weinig zorgen. Zolang het aantal beschikbare plaatsen klein is in verhouding tot het aanbod en zolang bovendien het materiaal een vrij groot percentage geschikten bevat, doet het in feite ook niet zo heel veel terzake, of wellicht een bepaald onderdeel voor de voorselectieprocedure in het geheel niet valide is: in het materiaal, dat in de eigenlijke selectiegang komt, bevinden zich dan toch nog wel voldoende geschikte kandidaten. Dit wordt uiteraard anders, wanneer men zich niet meer kan permitteren, om in het kader van de voorselectie een vrij belangrijk aantal geschikten te verspelen. Dan is het zaak, zich ook terdege met de validiteit van de voorselectie bezig te houden en het effect hiervan zo mogelijk kwantitatief door te rekenen. Dit kan geschieden op de wijze, die in het derde artikel is aangegeven, bij de bespreking van het gebruik van de gegevens van TAYLOR en RUSSELL: in feite zal men bij een selectie in enige trappen dezelfde redenering enige malen achtereenvolgens toepassen. Bij een gegeven toevalskans in het uitgangsmateriaal, een selectieprocedure met een bepaalde validiteit en een zekere selectieverhouding, krijgt men aan het einde van de eerste trap van het onderzoek een voorgeselecteerde groep, waarvan het percentage geschikten — of althans de verwachting dienaangaande — bekend is. Op deze groep kan men de tweede trap van de selectieprocedure toepassen en ook het effect hiervan kan worden berekend. Ook de invloed op de kostprijs kan worden berekend, waarbij de recente ontwikkeling ook weer wordt gekenmerkt door het streven, te werken met modellen, die mathematisch beschreven kunnen worden en die zich daardoor lenen voor kwantitatieve analyse.

Belangrijker nog dan het voorgaande is wellicht de toepassing van de sequente selectie op individuele keuzeproblemen, bijvoorbeeld het nemen van een beslissing inzake de keuze tussen v.h.m.o. en u.l.o. Gesteld de onderzoekresultaten voeren tot de constatering, dat iemand een weliswaar niet zeer grote, maar toch ook niet te verwaarlozen kans van slagen op het v.h.m.o. heeft — bijvoorbeeld in de orde van grootte van 20% (dit correspondeert met een IQ van circa 116 in afbeelding 10). De verwachting t.a.v. het economisch rendement geeft ons in dit geval niet veel houvast: volgens afbeelding 11 zou u.l.o. iets te prefereren zijn, maar deze figuur berust op geschatte gegevens en een kleine wijziging in de getalverhoudingen zou reeds tot een omkering van de situatie kunnen leiden. Er is dus in ieder geval geen reden, om de keuze van v.h.m.o. met klem te ontraden, anderzijds moet echter serieus rekening worden gehouden met de mogelijkheid van een zeer teleurstellende ontwikkeling. Volgens het principe van sequente selectie kan men nu een voorlopige beslissing nemen, bijvoorbeeld v.h.m.o. kiezen, maar met de aanwijzing, dat de eerste indicatie, dat de aanvankelijke ontwikkeling niet bevredigend verloopt, voldoende moet zijn om over te gaan op u.l.o. In dit geval is dan een niet-bevorderd worden in de eerste klas, of zelfs een eerdere aanwijzing, dat geen bevordering zal plaats vinden, voldoende om een dergelijke beslissing te nemen.

In een ander geval, waarin de kans van slagen groter is, behoeft een falen

in de eerste klas nog niet te worden opgevat als een voldoende indicatie om de plannen te wijzigen. In dit geval heeft iemand, die in de eerste klas struikelt, een nog lang niet te verwaarlozen kans op het behalen van het diploma. Nu ligt het voor de hand, om de beleidslijn zo te formuleren, dat men de situatie nog eens aanziet: een bevordering na een doublure in de eerste klas

IQ	bij overgang van lager naar voortgezet onderwijs	bij niet-bevordering na 1 jaar	bij niet-bevordering in II na doublure in I	Algemene strategieregels
110	geen v.h.m.o. kiezen			1. Bij bevordering studie voortzetten
115		v.h.m.o. verlaten		2. Bij niet-bevordering, als voorgaande klas vlot doorlopen is, doubleren
120	v.h.m.o. kiezen		v.h.m.o. verlaten	3. Bij slechte studieresultaten, bij herhaling van een klas of nadat voorgaande klas gedoubleerd werd, school tussentijds verlaten
125		doubleren		
130			doubleren	

STRATEGIETABEL

Afbeelding 12

ligt in de lijn der verwachting en geeft dus geen aanleiding tot een speciale stellingname, maar indien vervolgens reeds de aanvankelijke ontwikkeling in de tweede klas reden tot bezorgdheid geeft, is er voldoende reden een overplaatsing naar een eenvoudiger type van onderwijs te overwegen.

In het algemeen zal men een dergelijk advies natuurlijk niet uitsluitend op de uitkomsten van intelligentieonderzoek baseren. Maar indien wij hier eenvoudigheidshalve toch het IQ als uitgangspunt voor het vastleggen van een beleidslijn gebruiken, zou deze in termen van sequente selectie in de volgende strategietabel kunnen worden vastgelegd. (Eenvoudigheidshalve zijn hierin de grenslijnen bij 'ronde getallen' gelegd; in de praktijk zouden zij nauwkeuriger moeten worden bepaald.)

Distribuerende selectie

De kwantificerende benaderingswijze, die in het voorgaande werd toegelicht aan het voorbeeld van de differentiatie tussen opleidingsmogelijkheden, is ook bij uitstek geschikt voor toepassing in het kader van distribuerende selectie. De vraagstelling is dan, uit een gegeven aanbod van kandidaten, een aantal functies, ieder met een bepaald aantal plaatsen, op optimale wijze te

bezetten. Voor elk van deze functies beschikken we over een predictor, waarbij in de eerste benadering verondersteld wordt, dat de economische waarde van de plaatsing van een kandidaat evenredig is aan de score, die hij op de predictor heeft behaald. Men moet nu de kandidaten zodanig over de functies verdelen, dat de som van de aan hen toegekende economische waarden een maximum is. Bij een enigszins aanzienlijk aantal functies en plaatsen, bijvoorbeeld wanneer een grote onderneming in het kader van een wervingscampagne tientallen of honderden werknemers wil aantrekken, is dit een verre van eenvoudige opgave. Het vaststellen van de economische waarde van een groot aantal functies, afhankelijk van de predictorscores, is op zichzelf reeds een zeer omvangrijke arbeid — hoewel deze opgave bij de huidige stand van de wetenschap in vele gevallen in principe wel op te lossen is. Voor het berekenen van een verdeling, die een maximaal rendement geeft, zal men dan echter zeker moeten beschikken over een computer van aanzienlijk vermogen. Dit geldt in het bijzonder, wanneer ook nog rekening moet worden gehouden met aanvullende gegevens, bijvoorbeeld verschillen in de kosten van plaatsing in de verschillende functies, eisen ten aanzien van het minimum aantal personen, dat in bepaalde functies moet worden geplaatst, en dergelijke. Is eenmaal echter de nodige voorarbeid verricht, dan zal door toepassing van 'operations research' althans in grote bedrijven het rendement van de selectieprocedure naar alle waarschijnlijkheid aanzienlijk kunnen worden verbeterd.

HOOFDSTUK V

VALIDITEIT

Inleiding

Met de voorgaande beschouwing zijn wij gegaan op een terrein, dat vooral in de toekomst voor de praktijk van de selectie betekenis zal kunnen krijgen. Thans echter keren wij terug naar de hedendaagse realiteit, omdat afsluitend nog enkele consequenties van selectie moeten worden gezien. Het eerste probleem, dat ons zal bezig houden, is, in hoeverre selectie eventueel zijn eigen resultaten vernietigt, doordat na de selectie de beoordeling op andere wijze plaats vindt dan te voren.

De invloed van selectie op de normstelling

Het is een bekend feit, dat de menselijke beoordelingswijze, in termen van goed, middelmatig, en slecht, als regel afgestemd is op datgene, wat men redelijkerwijze aan prestaties meent te mogen verlangen, dus op het gemiddelde prestatieniveau in de te beoordelen groep. Gaat men echter een nieuwe, valide selectiemethode invoeren, dan worden vooral de zwakkere krachten geëlimineerd, zodat het gemiddelde prestatieniveau in de groep hoger wordt. Na verloop van tijd zal de beoordelaar zijn normen hieraan gaan aanpassen, zodat het aantal personen, dat als slecht (onvoldoende) wordt beoordeeld, weer gaat toenemen. Een dergelijk verschijnsel is bij het onderwijs inderdaad geconstateerd. Volgens de zogenaamde wet van POSTHUMUS is er een algemene tendentie, bij de beoordeling van een groep ongeveer de getalverhouding 1 : 2 : 1 aan te houden, bij de indeling in goeden, middelmatigen en slechten. Deze verhouding wordt niet door iedereen en te allen tijde gehanteerd. Maar wanneer men grote aantallen beoordelingen sommeert, komt de genoemde verhouding zeer duidelijk naar voren. Bij de analyse van schoolcijfers van een h.b.s. over een periode van bijna een eeuw — dwz. een periode waarin door veranderingen in de sociale situatie ongetwijfeld belangrijke veranderingen in de begaafdheidskarakteristiek van de leerlingen zijn opgetreden — werd door onderzoekers van het Nutsseminarium voor Paedagogiek te Amsterdam gevonden, dat inderdaad het percentage onvoldoendes binnen zeer enge grenzen door de jaren heen constant bleef. Dit komt enerzijds, omdat men zijn beoordeling aanpast aan het gemiddeld gehalte van het leerlingenmateriaal, terwijl anderzijds ook bij de diepgang en de uitvoerigheid, waarmee de leerstof wordt behandeld, onbewust rekening wordt gehouden met de capaciteit van de groep.

In bedrijfssituaties zal dikwijls een soortgelijk effect optreden. Er zijn hier echter wel uitzonderingsgevallen, waar een uitwendige, absolute maatstaf beschikbaar is, om de prestaties te beoordelen, bijvoorbeeld doordat een productiemachine het tempo van de arbeid regelt. In zeer veel gevallen ontbreekt echter de uitwendige maatstaf, waardoor men evenzeer als in het onderwijs genoodzaakt is, zich te baseren op de subjectieve schatting, die uitgaat van

de redelijk te achten prestatie. Zelfs als door middel van arbeidsanalyse ogenschijnlijk wel een mogelijkheid tot absolute normstelling kan worden verkregen, zal men gewoonlijk kunnen opmerken, dat na verloop van tijd, bijvoorbeeld na een wijziging in het arbeidsproces, toch een herziening van de normen plaats vindt.

Het voorgaande houdt dus in, dat in feite de achterhoede een bescherming vormt voor de middelmatige werkers. Elimineert men een gedeelte der slechten, dan wordt met een deel der middelmatigen een nieuwe achterhoede gevormd, die na enige tijd weer als ongeschikt wordt beoordeeld.

Dit verklaart, waarom men dikwijls meent te kunnen constateren, dat een nieuwe, verbeterde selectieprocedure, na verloop van tijd toch niet de resultaten oplevert, die men er op grond van de eerste indrukken van meende te mogen verwachten. Natuurlijk kan het ook wel eens voorkomen, dat de aanvankelijke successen op toevalstreffers berusten, maar dikwijls zal toch de onbewuste plaats vindende normherziening de reden zijn, waarom men na enige tijd weer evenveel ongeschikten aantreft, als tevoren.

Uiteraard pleit het voorgaande niet tegen selectie. Immers, er is toch een positief resultaat geboekt, want de volgens de nieuwe regels geselecteerde groep is van gemiddeld beter niveau dan de oude en daardoor in staat om een kwalitatief en kwantitatief betere prestatie te leveren.

Een ander aspect van dit verschijnsel is, dat predictoren hun werkzaamheid geheel of gedeeltelijk verliezen. In het verleden is bijvoorbeeld herhaaldelijk geconstateerd, dat de kans van slagen in het v.h.m.o. in de eerste plaats een functie is van de intelligentie. Gaat men echter de v.h.m.o.-candidaten op intelligentie selecteren, dan wordt de groep leerlingen ten aanzien van dit kenmerk homogener en dit heeft, zoals ook in het vorige artikel gedemonstreerd, tot gevolg, dat de correlatie tussen intelligentie en schoolsucces geringer wordt. Indien echter door de normherziening toch het percentage afgewezenen hetzelfde blijft, betekent dit, dat andere factoren — bijvoorbeeld ijver, stabiliteit, interesse, aanpassing aan de klassegroep en dergelijke — meer verband gaan houden met het schoolresultaat. Zou men vervolgens ook op dergelijke kenmerken, voorzover zij van min of meer constante aard zijn, gaan selecteren, dan worden ook deze predictoren grotendeels uitgeschakeld en het schoolsucces zal dan overwegend afhankelijk worden van invloeden van toevallige aard.

Het is dus denkbaar, dat door een voortgezette selectie een punt wordt bereikt, waarop verdere selectie niet mogelijk is, terwijl toch het percentage ongeschikten even groot is als vroeger. Dit punt is in ons land in de praktijk nog lang niet bereikt. Wel echter kan men momenteel reeds duidelijk constateren, dat het percentage leerlingen met onvoldoende prestaties, bij wie tevens een relatief laag intelligentiepeil wordt gevonden, aanzienlijk geringer is, dan bijvoorbeeld kort na de oorlog, terwijl daarentegen het percentage uitvallers met stoornissen in de werkhouding en dergelijke toeneemt.

Validiteit

Men kan het voorgaande verschijnsel ook van een andere zijde bezien. Wij hebben namelijk tot dusver uitsluitend gelet op de groep, die het selectieproces heeft doorstaan. Voor een objectieve beoordeling van de validiteit van de selectie dient men echter ook te letten op de groep, die bij selectie uitvalt. Vaak onttrekt deze groep zich aan onze beoordeling, wat betreft de aantallen geschikten en ongeschikten onder hen. Maar er zijn gevallen, waarin we over dit punt wel geïnformeerd kunnen zijn, bijvoorbeeld doordat gedurende een zekere tijd de selectiegegevens verzameld zijn, zonder dat hieraan praktische consequenties werden verbonden. Op grond hiervan kan men zeggen – wat ook langs theoretische weg kan worden afgeleid – dat de algemene regel is, dat we een gunstiger beeld van de validiteit van de selectieprocedure krijgen, indien we niet alleen met de geplaatsten, maar ook met de niet-geplaatsten rekening houden.

Treedt na de invoering van de selectie een normverschuiving op, waardoor binnen de groep geaccepteerden de validiteit van de predictor afneemt, dan geldt, dat complementair de validiteit van de predictor ten aanzien van de niet-geaccepteerden toeneemt.

Praktische evaluatie van de validiteit

De beoordeling van de praktische waarde, die aan een predictor met een bepaalde validiteit moet worden toegekend, is ook een probleem, dat de aandacht verdient. De eigenaardigheden van een statistische grootheid, zoals een correlatiecoëfficiënt is, maken, dat de wijze van interpretatie zorgvuldig overwogen moet worden. In de eerste plaats dient rekening te worden gehouden met de frequentieverdeling, zowel van de predictorscores, als van de criteriumwaarden. De meeste interpretatiewijzen stellen aan deze verdeling bepaalde voorwaarden, die men in leerboeken over statistiek nader beschreven vindt. In de praktijk kan men veelal er van uitgaan, dat, indien de beide randverdelingen niet te veel van de normale Gauss'se verdeling afwijken, wel bij benadering aan deze voorwaarden wordt voldaan.

Vaak drukt men dan de effectiviteit van de predictor uit in de mate van voorspelbare variantie. D.w.z. men vergelijkt de afwijkingen tussen de voorspelde en de werkelijke criteriumwaarden bij gebruik van de predictor, met die, welke op grond van een toevalsvoorspelling zouden optreden. In het tweede geval zal men een grotere mate van spreiding in de afwijkingen vinden, de variantie is een statistische maatstaf voor deze spreiding. Men kan nu aangeven, met welk percentage de toevalsvariantie kan worden gereduceerd, door toepassing van de predictor. Dit percentage blijkt evenredig te zijn met het kwadraat van de validiteit. Dus bij een validiteit $r = .40$ wordt de variantie met 16% gereduceerd, bij $r = .50$ met 25% etc. Bij een volmaakte correlatie $r = 1.00$ is de reductie 100%, dan zijn er geen verschillen meer tussen de voorspelde en de werkelijke criteriumwaarde. In het licht van deze interpre-

tatiwijze schijnt de waarde van een predictor met een middelmatige validiteit niet zeer groot. De resterende variantie is immers $(1-r^2)$ maal de toevalsvariantie. Pas bij een $r > .70$ is de variantie dus op de helft gereduceerd.

Het werken met het begrip variantie heeft een nadeel, omdat dit begrip moeilijk concreet kan worden voorgesteld. Wil men een indruk krijgen van de mate, waarmee de orde van grootte van de gemaakte fouten gemiddeld verandert, dan doet men beter een andere index te gebruiken, die gebaseerd is op de vierkantswortel uit de variantie. Het percentage, waarmee de gemiddelde orde van grootte van de fouten wordt gereduceerd, wordt dan aangegeven door de formule $(1 - \sqrt{1 - r^2}) \times 100$. Bij een test met een validiteit van $r = .60$, levert deze formule als uitkomst, dat de bij de voorspelling gemaakte fouten in doorsnee 20% kleiner zullen zijn, dan zonder toepassing van deze voorspeller. Op deze wijze bezien is de waarde van de predictor zo mogelijk nog minder imponerend, dan wanneer we zeggen, dat dezelfde test een reductie van 36% van de toevalsvariantie mogelijk maakt. En we hebben reeds gezien, dat een validiteit van $r = .60$ dicht bij de bovengrens ligt, van datgene, wat we in de praktijk gewoonlijk mogen verwachten.

Het beeld wordt echter gunstiger, wanneer we letten op het rendement van de selectie. Dit rendement is, behalve van de validiteit van de predictor (r), afhankelijk van verschillende factoren.

Hiertoe behoort in de eerste plaats de potentiële selectiewinst, dwz. het verschil in economische waarde tussen een goed geschikte en een weinig geschikte functionaris.

Hoe groter dit verschil, des te groter natuurlijk ook het mogelijk rendement van een selectieprocedure. In de a priori gegeven groep kandidaten heeft iedere gegadigde een bepaalde economische waarde, er is dus ook een a priori verdeling van de economische waarden, met een bepaalde spreiding (s). Het rendement is ook afhankelijk van de selectieverhouding of de cut-off score. Als de scores op de predictor bij benadering een normale verdeling vertonen, is het verband tussen cut-off score (c) en selectieverhouding bekend, zodat men eerstgenoemde als maatstaf kan gebruiken, hetgeen in dit geval voordelen heeft.

Voorts is het rendement natuurlijk ook afhankelijk van de kosten (K) van de selectie. Indien men nu veronderstelt, dat de economische waarde van een kandidaat recht evenredig is met zijn predictorscore, dan kan men langs wiskundige weg bewijzen, dat het rendement van de selectie per man wordt gegeven door de relatie

$$\text{Rendement} = r.s.f(c) - K$$

die dus recht evenredig met de validiteit is. In deze relatie stelt $f(c)$ een bepaalde functie van c voor.*)

*) Namelijk de waarde van de ordinaat van de normale verdeling, op het punt c . Aangezien deze waarde maximaal is bij de mediaan, volgt hieruit, dat het economisch rendement van de selectie maximaal is bij een selectieverhouding van .50, althans indien de veronderstellingen die aan het model ten grondslag zijn gelegd, geldig zijn. (Vgl. ook hoofdstuk III voor een ander model met andere consequenties.)

In economisch opzicht heeft men dus bij gebruik van een predictor met een validiteit $r = .60$ reeds 60% van de winst bereikt, die theoretisch mogelijk is. (Althans indien k relatief klein is ten opzichte van s .) Dit geeft uiteraard een heel wat gunstiger beeld van de waarde van de selectieprocedure.

Dit neemt niet weg, dat het natuurlijk toch wenselijk is, te streven naar verhoging van het rendement. Daarbij zal dan echter wel in de eerste plaats moeten worden gelet op de betrouwbaarheid van de beoordeling achteraf. Wij hebben in het eerste hoofdstuk reeds gezien, dat de situatie hier in de praktijk nog zeer veel te wensen overlaat, zowel door onzuiverheden in de normatieve uitgangspunten, waardoor verschillende criteria door elkaar worden gebruikt, als door gebreken in de mogelijkheden om de beoordeling objectief te doen geschieden en de daaruit voortvloeiende geringe mate van overeenstemming tussen de beoordelaars. Langs theoretische weg kan men aantonen, dat er een bovengrens is voor de validiteit van een selectieprocedure, die gegeven wordt door de vierkantswortel uit de betrouwbaarheid van de beoordeling achteraf. Heeft deze bijvoorbeeld een betrouwbaarheid, die uitgedrukt wordt door een correlatie van $.64$ tussen de uitspraken van de verschillende beoordelaars (een waarde, die in de praktijk niet eens altijd wordt bereikt), dan is het theoretisch maximum voor de validiteit van de selectieprocedure gelijk $r = .80$. Met een predictor, die een validiteit van $r = .60$ heeft, zijn we dit maximum dus reeds betrekkelijk dicht genaderd, vooral als men in aanmerking neemt, dat er altijd een marge zal blijven voor de werking van 'toevalsfactoren'. Verbeterd men echter de betrouwbaarheid van de beoordeling, dan komt er meer ruimte voor verdere arbeid aan de verbetering van de selectie. Ook hier ligt dus een belangrijke opgave voor de verdere ontwikkeling van de selectiewetenschap.

LITERATUUR

(De literatuuropgaven zijn gerangschikt naar de volgorde van bespreking der onderwerpen, waarop zij betrekking hebben. De cijfers verwijzen naar de hoofdstukken.)

1. TIFFIN, J. *Industrial psychology*, London 1951.
 BLUM, M. L. *Industrial psychology and its social foundations*. New York 1956.
 DE GROOT, A. D. *Methodologie*. Den Haag 1961.
2. (*American Psychological Association*) *Technical recommendations for psychological tests and diagnostic techniques*. Washington 1954.
 CRONBACH, L. J. *Essentials of psychological testing*. New York 1960.
 GUILFORD, J. P. *Psychometric methods*. New York 1954.
 LIENERT, G. A. *Testaufbau und Testanalyse*. Weinheim 1961.
 WIEGERSMA, S. Psychometrische betrouwbaarheid. *Ned. Tijdschrift Psychol.* 15 (1960) 22–46.
 V. D. HOEVE, A. H. KOHNSTAMM, PH., VAN VEEN, G. Stilleesstof als denkmateriaal en denkmaatstaf. (Med. Nutssemin. voor Paedagogiek nr. 24), Groningen, 1935.
 (*Centraal Bureau voor de Statistiek*) *Statistiek van het Hoger Onderwijs; studie-rendement van de kandidaatsstudie*. Zeist 1958.
 NIEUWENHUIS, H. De schoolloopbaan van ruim 500 u.l.o.-leerlingen. *Paed. Studiën* 31 (1954) 137–156.
 BARNETTE, W. L. Occupational aptitude patterns of selected groups of counseled veterans. *Psychol. Monographs* nr. 322 (1951).
 MATTHIJSEN, M. A. J. M., SONNEMANS, G. J. M. Schoolkeuze en schoolsucces bij v.h.m.o. en u.l.o. in Noord-Brabant. Tilburg 1959.
 (*Technische Hogeschool*) *Mislukking en vertraging in de studie*. Delft 1959.
3. WIEGERSMA, S. *Psychologie der beroepskeuze*. Utrecht 1961.
 MEEHL, P. E., ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. *Psychol. Bull.* 52 (1955) 194–216.
 TAYLOR, H. C., TUSSELL, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection. *J. Appl. Psychol.* 23 (1939) 565–578.
 HUISKAMP, J. De rentabiliteit van het psychologisch onderzoek ten dienste van het bedrijfsleven. Publicatie nr. 4 van de Ned. Ver. voor Bedrijfspsychologie, 1949.
4. CRONBACH, L. J., GLEESER, G. C. *Psychological tests and personnel decisions*. Urbana 1957.
 VAN NAERSSEN, R. F. *Selectie van chauffeurs*. Groningen 1963.
5. POSTHUMUS, K. *Levensgeheel en school*. Den Haag z.j.
 BROUWER, W. H. e.a. *Selectie en schoolsucces* (Med. Nutssem. voor Paedagogiek nr. 50). Groningen 1951.
 GUILFORD, J. P. *Fundamental statistics in psychology and education*. New York 1956.
 GULLIKSEN, H. *Theory of mental tests*. New York 1950.
 CRONBACH, L. J., GLEESER, G. C. als boven.

SUMMARY

In a series of five chapters the author discusses a number of problems related to selection. Both personnel selection and educational and vocational guidance present problems that come under this heading.

In the first chapter the four main components of any selection are briefly discussed: the person, the situation, the criterion and the way of judging persons and situations according to a criterion. Much stress is laid upon the problems that result when — unknown to each other — individual judges apply different criteria to the same case. Attention is also given to the ways in which a criterion can be made operational and to problems of validity.

The second chapter is concerned with predictors. The relation between any predictor and occupational or educational achievement is always a probability function. The nature of this function is demonstrated in a number of examples, e.g. the relation between school marks and success in a university course or between IQ and secondary school achievement. It is also shown that in the case of a combination of predictors the relation with success is again a probability function.

The theme of the third chapter is the efficiency of selection. Starting from the probability function that gives the relationship between IQ and grammar school achievement, the efficiency of using an IQ for grammar school selection is investigated. The effects of using different cut-off scores are examined in a numerical example.

Next it is stated that the relationship between a predictor and achievement can also be given by a correlation-coefficient — and the importance of antecedent probability and selection-ratio are stressed. Some of the Taylor-Russel data are presented in diagrams and from this the practical limitations are shown with regard to the efficiency of selection. The economic utility of selection is also discussed in this connection.

The fourth chapter starts with a problem of distributive selection. The example used is that of a choice between two types of secondary education, both with a particular probability function, related to IQ and with a specific economic value. A utility diagram is derived showing at which IQ-value the best cut-off score for making a choice between the two types can be placed. This leads to a general discussion of problems of operational research and from there to the advantages of sequential selection. The concept of a strategy of choices is elaborated.

The final chapter deals with a number of other problems of selection. One of them is that judges gradually adjust their norms to the average level of the group to be judged. In both industrial and in educational selection this gradually invalidates the effects of selection. Therefore, the concept of validity has to be reinterpreted. The chapter closes with a number of remarks on the effectiveness of selection following from modern decision theory.