

Verhandelingen

INTERESSE EN SCHOOLKEUZE

Belangstellingsonderzoek bij de differentiatie na de Lagere school

Dr. S. Wiegersma

Nederlands
Instituut
voor
Praeventieve
Geneeskunde



ERRATUM

De op pagina 37 vermelde formule voor de toevalsrelatie moet liden:

$$r = - m/108$$

De verwachte waarden voor de toevalsrelaties tussen de schalen variëren derhalve tussen $-.083$ en $-.167$.

u g c B

W 6 2 (2)

INTERESSE EN SCHOOLKEUZE

**BELANGSTELLINGSONDERZOEK BIJ DE DIFFERENTIATIE
NA DE LAGERE SCHOOL**

DRUK: N.V. v.h. BATTELJEE EN TERPSTRA — LEIDEN

**Uit de afdeling Geestelijke Gezondheid van het Nederlands Instituut voor
Praeventieve Geneeskunde te Leiden. Hoofd Prof. Dr J. Koekebakker.**

VERHANDELING VAN HET
NEDERLANDS INSTITUUT VOOR PRAEVENTIEVE GENEESKUNDE
XLIV

47CB
W 62
2)

INTERESSE EN SCHOOLKEUZE

BELANGSTELLINGSONDERZOEK
BIJ DE DIFFERENTIATIE
NA DE LAGERE SCHOOL

DOOR

Dr. S. WIEGERSMA

1959

VOORWOORD

Een onderzoek, zoals in deze verhandeling gerapporteerd, komt niet tot stand zonder team-werk. Daarom wil ik ook zeer gaarne mijn erkentelijkheid uitspreken jegens al degenen, die met mij aan dit onderzoek hebben gewerkt. Deze dank geldt in de eerste plaats de heer A. H. BOON VAN OSTADE psych. drs., die bij de coördinatie van veldwerk en materiaalverwerking mij van veel praktische zorgen heeft ontlast en bovendien een belangrijk aandeel heeft gehad in de statistische toetsing van de resultaten. Daarnaast dank ik ook de medewerkers, die bij de voorstudiën, veldwerk, de administratie der testuitkomsten of het rekenwerk hun steun hebben verleend: Mejuffrouw I. DOMMERHOLT, Mevrouw A. M. MEULEMANS-TRAVERSIN lic.psych., Mejuffrouw B. J. RAVENSWAAY, Mevrouw M. S. SALOMÉ-VOORHAGEN en de heren F. J. KELLENBACH, H. M. MELIËZER psych.drs., Mr. E. J. P. ROBERTS en J. P. WAMS. Voorts al de hoofden en onderwijzers van lagere scholen, die hun toestemming tot het onderzoek gaven of in hun klas als proefleider fungeerden en ook de directeuren en rectoren van scholen voor voortgezet onderwijs, die eveneens door hun welwillendheid de praktische verwerkelijking van mijn onderzoekplannen mogelijk maakten.

Erkentelijk ben ik ook de heren Ch. A. G. NASS en H. DE JONGE, voor hun advies bij een aantal statistische moeilijkheden, de heer N. H. DE KLEYN voor de verzorging van het tekenwerk en mejuffrouw R. H. F. DE RU voor haar bereidheid, de „Summary” op zijn taalkundige merites te beoordelen.

Mevrouw J. M. J. VAN WIJK-BOOM dank ik zeer hartelijk voor de bijzondere zorg en snelheid, waarmee zij het uittypen van het manuscript heeft verzorgd.

Leiden, mei 1959.

S. WIEGERSMA

INHOUDSOPGAVE

INLEIDING	1
Hoofdstuk I. ENKELE ALGEMENE BESCHOUWINGEN OVER HET INTERESSEBEGRIIP	
1.1. Begripsbepaling	7
1.2. Karakteristiek der interessen	9
1.3. Stabiliteit der interessen	13
Hoofdstuk II. INTERESSENONDERZOEK	17
2.1. Historisch overzicht	17
2.2. Hoofdvormen van tests	20
2.3. De studie van IRLE	30
Hoofdstuk III. DE BEROEPENINTERESSE TEST	43
3.1. Beschrijving van de test	43
3.2. De vertaling van de B I T	44
3.3. Handleiding voor de B I T	51
Hoofdstuk IV. NIVEAUVERSCHILLEN IN DE BELANGSTEL- LING	55
4.1. Probleemstelling	55
4.2. Eerste experimentele opzet	57
4.3. Statistische bewerking	60
4.4. Niveauverschillen in de bestaande schalen	62
4.5. Schaalconstructie	65
4.6. Tweede experimentele opzet	74
4.7. Intelligentie en interesse	81
Hoofdstuk V. DE PSYCHOMETRISCHE KWALITEITEN VAN DE B I T-SCHALEN	83
5.1. Vraagstellingen en experimentele opzet	83
5.2. Homogeniteitsanalyse	84
5.3. Revisie van de HA-schaal	91
5.4. Intercorrelaties	93
5.5. Betrouwbaarheid	94

Hoofdstuk VI. DE IJ KING VAN DE B I T-SCHALEN	99
6.1. De experimentele opzet van de ijking	99
6.2. De representativiteit van de steekproef	103
6.3. De normering van de test	108
6.4. Significantie van niveauverschillen	111
6.5. Toevalsscores	113
Hoofdstuk VII. VAN INTERESSETEST TOT INTERESSE- KARAKTERISTIEK	115
7.1. Het validatieprobleem	115
7.2. Operationele definitie	117
7.3. De belangstellingskarakteristiek	121
Hoofdstuk VIII. SEXEVERSCHILLEN IN DE BELANGSTEL- LING	123
8.1. Vraagstelling	123
8.2. Congruente validatie van de sexeverschillen	124
8.3. Sexeverschillen in de B I T	125
Hoofdstuk IX. INTERESSE EN SCHOOLKEUZE	131
9.1. Inleiding	131
9.2. Bestemming v.h.m.o.	133
9.3. Bestemming m.u.l.o. of handelsdagschool	135
9.4. Nijverheidsonderwijs, v.g.l.o. en praktijk bij meisjes .	136
9.5. Nijverheidsonderwijs, v.g.l.o. en praktijk bij jongens .	139
9.6. Beroepskeuze in de agrarische sector	141
9.7. Leerlingen in een overgangsjaar	142
9.8. Samenvatting	143
Hoofdstuk X. BELANGSTELLING EN RICHTINGKEUZE IN HET NIJVERHEIDSONDERWIJS VOOR JONGENS	146
10.1. Probleemstelling	146
10.2. Bankwerkers	147
10.3. Instrumentmakers	148
10.4. Electromonteurs	149
10.5. Loodgieters	149
10.6. Schilders	151
10.7. Straatmakers	152
10.8. Houtbewerkers	153
10.9. Bakkers	154
10.10. Typografen	155
10.11. Leeftijdsinvloeden	157

Hoofdstuk XI. PREDICTIEVE VALIDITEIT	160
11.1. Probleemstelling	160
11.2. Experimentele opzet	164
11.3. Een contrôle op de ijkingsgroep	165
11.4. Schoolsucces in het eerste jaar in het v.h.m.o.	167
11.5. De invloed van strenge selectie	169
11.6. De waarde van een „gewogen” schaal	172

Hoofdstuk XII. PERSOONSKENMERKEN EN B I T-SCORES .	175
12.1. Vraagstelling	175
12.2. Leeftijdsverschillen in het ijkingsmateriaal	175
12.3. Woongebied	176
12.4. Beroepsniveau vader	177
12.5. Beroepssector vader	178
12.6. Onderwijsrichting	179

BIJLAGEN

I. STATISTISCHE SIGNIFICANTIE VAN VOORKEURSVER- SCHILLEN	186
II. DETERMINEERTABEL VOOR DE RICHTINGKEUZE IN HET VOORTGEZET ONDERWIJS	188
III. GRAFISCHE DETERMINEERTABEL VOOR HET VOORTGEZET ONDERWIJS VOOR JONGENS	190
IV. GRAFISCHE DETERMINEERTABEL VOOR HET VOORTGEZET ONDERWIJS VOOR MEISJES	191
V. BEROEPENINTERESSETEST	192
LITERATUURLIJST	194
SAMENVATTING	203
SUMMARY	207

INLEIDING

In the wise choice of a vocation there are three broad factors: (1) a clear understanding of yourself, your aptitudes, abilities, *interests*, ambitions, resources, limitations and their causes; (2) a knowledge of the requirements and conditions of success, advantages and disadvantages, compensation, opportunities and prospects in different lines of work; (3) true reasoning on the relations of these two groups of facts. FR. PARSONS (104).

De betekenis van de belangstelling in verband met de keuze van school, vakopleiding en beroep wordt algemeen erkend. Het ligt voor de hand, dat iemand, die taken te verrichten krijgt, welke zijn interesse hebben, daarvoor zijn werk met meer plezier zal verrichten. Hij zal waarschijnlijk eerder doorzetten als er moeilijkheden optreden, hij zal meer geneigd zijn tot spontane en extra activiteit. Kortom hij zal beter gemotiveerd zijn dan iemand, die onder overigens gelijke omstandigheden, dezelfde taken oninteressant vindt. Men mag ook verwachten, dat bij een gegeven begaafdheidsstructuur, werkhouding, sociaal milieu, gezondheidstoestand etc., de kans op succes bij de keuze van een bepaalde school- of beroepsrichting zal toenemen, naarmate deze richting beter past bij de interesses van de candidaat.

Daarom zal ook in het kader van voorlichting bij school- en beroepskeuze rekening moeten worden gehouden met de belangstelling van de adviesvrager. Dat zal de leek doen, wanneer hij eens dergelijke raad geeft, dat kan ook de professionele adviseur niet anders doen. Het hierboven weergegeven citaat van Frank PARSONS, de grondlegger van de wetenschappelijke beroepskeuzevoorlichting, toont aan, dat de betekenis van de belangstelling in dit verband reeds aanstonds is onderkend. Het behoort dus zo te zijn, dat de beroepskeuze-adviseur tracht, zich zo systematisch mogelijk te oriënteren over de belangstelling van de adviesvrager. Hij zal zich daarbij moeten richten op die aspecten van de belangstelling, welke als motiverende kracht bij de ontwikkeling op school of in het beroep de grootste praktische betekenis hebben.

In de voorlichting bij school- en beroepskeuze wordt in ons land als regel gebruik gemaakt van de uitkomsten van een psychologisch onderzoek. Het zou in dit verband dus voor de hand liggen, dat men zich ook over de aard en de sterkteverhoudingen van de interesses met behulp van een test oriënteert. De realiteit in Nederland beantwoordt echter niet aan de verwachting, die in de voorgaande zin werd uitgesproken. Terwijl ten

aanzien van verschillende onderdelen van de begaafdheidsstructuur en ook ten aanzien van allerlei aspecten van de persoonlijkheidskarakteristiek wél een, soms diepgaand psychologisch onderzoek wordt ingesteld, geschiedt dit ten aanzien van de belangstelling niet. Zelfs de middelen daartoe zijn in ons land, in tegenstelling tot vele andere landen, nog niet aanwezig. Wij beschikken nog niet over een deugdelijk geijkte en gevalideerde belangstellingstest, althans is een dergelijke test voor de practizerende psycholoog niet algemeen beschikbaar. Deze moet zich daarom tevreden stellen met een poging, in een interview een analyse van de interessesstructuur van de adviesvragende te maken, waarbij hij zich baseert op de reacties van de gesprekspartner op allerlei gesuggereerde activiteiten en situaties en voorts op diens liefhebberijen en wijze van vrije-tijdsbesteding. De op deze wijze verkregen inzichten kan hij trachten aan te vullen met behulp van de gegevens ontleend aan een zogenaamde beroepenlijst. Dit laatste hulpmiddel, waarover in een later hoofdstuk nog zal worden gesproken, is echter van zeer twijfelachtige waarde. Daarbij komt nog, dat wij over de leeftijds-karakteristiek van de interesses en ook over de wijze waarop de belangstelling samenhangt met allerlei andere factoren, nog slechts zeer weinig in detail weten.

In ons land vindt een eerste grote differentiatie in het kader van de school- en beroepskeuze plaats aan het einde van de lagere school. Tevoren hebben alle kinderen in principe ongeveer hetzelfde onderwijs genoten. Ongetwijfeld zijn er ook tussen verschillende typen lagere scholen verschillen, tussen een opleidingsschool, een „gewone” lagere school, een school met eenvoudig leerplan in een grote stad en een kleine dorpschool. Maar deze verschillen hebben slechts een vrij bescheiden betekenis, voornamelijk komen zij tot uiting in het tempo van het onderwijs. Indien nodig kan een leerling de achterstand, die hij op een bepaalde school heeft opgelopen, nog gemakkelijk compenseren, b.v. door een enkel jaar aanvullend onderwijs op een andere lagere school. Aan het einde van de lagere school vindt echter een richtingskeuze plaats, met veel verder strekkende consequenties. De mogelijkheden, waartussen een keuze moet worden gedaan, betekenen bovendien niet alleen een differentiatie naar de aard van het programma, maar ook naar het niveau van abstractie in het onderwijs en naar tempo. De verschillen die hier ontstaan zijn zo groot, dat een correctie van een verkeerde keuze al na betrekkelijk korte tijd zeer bezwaarlijk wordt. De wenselijkheid, dat de middelen ter beschikking staan, om bij de keuze van het voortgezet onderwijs zo nodig een effectieve voorlichting te geven, wordt daarmee reeds voldoende geargumenteed.

Ten aanzien van het belangstellingsonderzoek betekent dit dus, dat het in de Nederlandse situatie van belang is, speciaal over de leeftijdsfase omstreeks de overgang van lager naar voortgezet onderwijs, d.w.z. in hoofdzaak ten aanzien van de 12—14-jarigen, een nader inzicht in de structuur en de correlaten van de belangstelling te krijgen. Over de karak-

teristiek van deze leeftijdsgroep is tot dusver nog maar betrekkelijk weinig bekend. Het leeuwenaandeel der onderzoeken op het terrein van de belangstellingspsychologie en van de interressentests, heeft tot dusver in de Verenigde Staten plaatsgevonden. In het Amerikaanse onderwijsstelsel speelt echter de differentiatie aan het einde van de lagere school geen rol, zodat er daar geen aanleiding was, aan de leeftijdsgroep die voor ons land van betekenis is, bijzondere aandacht te schenken.

Het is daarom gewenst, de problematiek van het interessenonderzoek in ons land in studie te nemen. Uit het voorgaande volgt, dat daarbij in hoofdzaak twee problemen kunnen worden onderscheiden. In de eerste plaats bestaat er behoefte aan een deugdelijke interressentest, met name een test, die ook reeds aan het einde van de lagere school kan worden gebruikt. In de tweede plaats is het gewenst, een beter inzicht te krijgen in de karakteristiek van de belangstelling van de categorie kinderen, die op het punt staat, de lagere school te verlaten.

Met betrekking tot het eerstgenoemde probleem, het ter beschikking stellen van een belangstellingstest, rijst preliminair de vraag, of een bestaande buitenlandse methode voor Nederland moet worden bewerkt, of dat een geheel nieuwe test van eigen fabricaat moet worden ontwikkeld. Beide mogelijkheden hebben hun voor- en nadelen. Tegen alle bestaande methoden kunnen wel bepaalde bezwaren worden aangevoerd. Deze bezwaren kunnen betrekking hebben op de tijdsduur van de test, het feit dat een gedwongen keuze tussen bepaalde alternatieven moet worden gemaakt, de aard van de items of van de testcategorieën, de scoringsmethode, etc. Het zou verleidelijk zijn, te trachten, een eigen test te ontwikkelen, die vrij zou zijn van al deze tekortkomingen.

Het is echter gemakkelijker, een dergelijk ideaal te formuleren, dan het te verwezenlijken. De lange tijdsduur van de test blijkt een consequentie te zijn van het streven naar een goede betrouwbaarheid. De ingewikkelde scoringsmethode hangt samen met het verlangen, de validiteit zo hoog mogelijk op te voeren. Het aantal geschikte items is beperkt, zodat men min of meer noodgedwongen ook wel een aantal items, waartegen bezwaren zijn aan te voeren, moet opnemen. Het principe van de gedwongen keuze blijkt noodzakelijk, omdat aan een vrije keuze nog veel groter methodologische bezwaren zijn verbonden.

Dit wil nog niet zeggen, dat niet de ene test methodologisch en praktisch van beter kwaliteit zou kunnen zijn dan de andere; maar ook de beste tests blijken hun aanvechtbare kanten te hebben. De realiteitszin gebiedt te beseffen, dat een eigen test, hoe goed ook geconstrueerd, toch zeker ook wel bepaalde gebreken zal vertonen. Onder deze omstandigheden moet de vraag worden gesteld, of men zich niet door het bewerken van een goede buitenlandse methode veel grondwerk kan besparen. Het ontwikkelen van een eigen test vergt een tijdrovende voorarbeid, waarin een groot aantal items op hun bruikbaarheid moeten worden onderzocht,

teneinde een voldoende aantal van goede kwaliteit te verzamelen, waaruit het definitieve model van de test kan worden samengesteld. Deze fase in de testconstructie kan grotendeels worden overgeslagen, indien een test wordt gebruikt met items, waarvan de deugdelijkheid elders reeds is gebleken. Indien een dergelijke test is ontwikkeld in een land dat ten aanzien van vele relevante kenmerken, i.c. economische structuur, onderwijs-systeem, beroepenverdeling en beroepenwaardering, op Nederland gelijk, mag men verwachten, dat de grote meerderheid van de items ook in ons land bruikbaar zal blijken te zijn. Bovendien kan men, door een test te bewerken die in het buitenland reeds het voorwerp van statistisch onderzoek is geweest, snel tot een methode komen, waarvan men met een redelijke mate van zekerheid mag verwachten, dat deze aan vrij hoog te stellen eisen van deugdelijkheid zal voldoen. Uiteraard moet deze deugdelijkheid nog wel proefondervindelijk worden aangetoond; maar de kans dat het desbetreffend onderzoek positieve resultaten oplevert, is gunstig. Tenslotte kan men de tijd, die men op de voorbereiding uitspaart, nuttig gebruiken, om te trachten voor eventuele tekortkomingen van de gekozen methode een verbetering te vinden.

Deze overwegingen hebben er toe geleid, dat besloten werd een reeds bestaande test voor Nederland te bewerken. Aanvankelijk heeft het voornemen bestaan, hiervoor de *Kuder Preference Record* te gebruiken, maar uiteindelijk is de keuze gevallen op de *Berufs-Interessen-Test* (B I T) van M. IRLE. Deze methode is nauw aan de KUDER-test verwant, maar heeft een iets andere vorm, waardoor de praktische verwerking van de testresultaten eenvoudiger wordt.

De B I T is door IRLE ontwikkeld ten behoeve van het onderzoek van personen, die de lagere school reeds ten minste enige jaren hebben verlaten. De meeste door hem onderzochte proefgroepen waren gemiddeld 16 à 17 jaar oud; in de situatie waarin hij werkte bestond er voor hem weinig aanleiding, aandacht te schenken aan een aanzienlijk jongere leeftijdsgroep. We hebben er echter reeds op gewezen, dat in ons land de school- en beroepskeuzeproblematiek vooral bij de 12- tot 14-jarigen van praktische betekenis is. Dit blijkt ook uit de gegevens, die het C.B.S. verstrekt ⁽³⁶⁾ over de in ons land verstrekte beroepskeuze-adviezen (de schoolkeuze daarbij inbegrepen). De laatste jaren heeft ruim 60% van deze adviezen betrekking gehad op personen, jonger dan 15 jaar en het aandeel van de voorlichting aan deze jeugdigen blijkt zelfs nog van jaar tot jaar toe te nemen. Daarom is de B I T speciaal voor deze leeftijdsgroep bewerkt, zowel de door ons verzamelde ijkingsgegevens als de statistische toetsing hebben hierop betrekking.

In het kader van de tweede hierboven genoemde vraagstelling, is de B I T vervolgens gebruikt als hulpmiddel om de belangstelling van de 12—14-jarigen en voornamelijk die van de leerlingen omstreeks het einde van de zesde klasse der lagere school nader te leren kennen. Een onder-

zoek met behulp van de test bij een aantal categorieën kinderen uitgevoerd, gaf de gelegenheid, de relatieve sterkte van de verschillende interesses te leren kennen en ook om de correlatie van de belangstelling met diverse uitwendige criteria te onderzoeken. Tendele kunnen dergelijke onderzoeken ook dienstbaar worden gemaakt aan de validatie van de test. In principe kan men de hier gevolgde werkwijze zien als een poging, de interessenstructuur van leerlingen omstreeks het einde van de lagere school te bestuderen op een wijze, die psychodiagnostisch hanteerbaar is. D.w.z. als een formulering van een aantal conclusies, betreffende deze interessenstructuur, in categorieën, ontleend aan de test.

Methodologisch beschouwd ⁽⁵⁵⁾ kan men aan de hier te rapporteren opzet drie aspecten onderscheiden. Het eerste gedeelte van de studie heeft een instrumenteel karakter, het beoogt een elders ontwikkeld hulpmiddel van het psychodiagnostisch onderzoek door bewerking geschikt te maken voor gebruik in de Nederlandse situatie. Een dergelijk instrument is gebaseerd op een aantal principes, enerzijds van psychometrische aard, anderzijds betrekking hebbende op bepaalde „constructs” ten aanzien van de interesses. De toetsing van deze principes in het kader van de test, is een tweede opgave van deze studie. Ten derde heeft de studie ook een exploratief karakter, er wordt met behulp van de test empirisch materiaal verzameld, dat bruikbaar kan worden gemaakt aan de hypothesevorming in het kader van de belangstellingspsychologie.

Ter introductie tot deze studie scheen het gewenst, eerst, zij het betrekkelijk kort, in te gaan op een aantal reeds bekende gegevens met betrekking tot de belangstellingspsychologie en het interessenonderzoek. Het eerste hoofdstuk zal worden gewijd aan een begripsbepaling, die wordt gevolgd door een discussie over de karakteristiek der interesses. Hierop aansluitend komt nog het voor de praktijk zeer belangrijke probleem van de constantie der interesses ter sprake. In het tweede hoofdstuk wordt ingegaan op de problematiek van het belangstellingsonderzoek. Na een beknopt historisch overzicht volgt eerst een bespreking van een aantal gezichtspunten, die in de methodiek van het belangstellingsonderzoek een rol spelen. Enkele hoofdvormen van tests worden daarna behandeld. Uitvoeriger wordt dan ingegaan op het werk, dat door IRLE, in het kader van de constructie van de B I T, reeds met het oog op de toetsing van de test is verricht. In deze beide hoofdstukken zal de bespreking niet uitsluitend worden beperkt tot het belangstellingsonderzoek bij 12—14-jarigen, maar zullen de belangstellingstests en hun problemen in het algemeen ter sprake komen. Dit is gewenst, omdat een groot deel van de problematiek vrijwel onafhankelijk is van de specifieke leeftijdscategorie, waarvoor de test wordt gebruikt.

In het derde hoofdstuk worden de test, zijn vertaling en de handleiding voor het gebruik van de test behandeld. Ook het vierde hoofdstuk draagt een instrumenteel karakter, hierin worden de problemen, samenhangende

met de constructie van enkele nieuwe belangstellingsschalen behandeld. De drie hierna volgende hoofdstukken hebben betrekking op de toetsing en ijking van de test. De verschillende pogingen, tot validering van bepaalde aspecten van de test, hebben ten dele ook een exploratief aspect. Geleidelijk gaat daardoor het tweede deel van de studie in het derde over. In de laatste vijf hoofdstukken, die een overwegend exploratief karakter dragen, worden de specifieke kenmerken van de belangstellingsstructuur van verschillende subgroepen uit de onderzochte leeftijdsgroep bestudeerd.

HOOFDSTUK I

ENKELE ALGEMENE BESCHOUWINGEN OVER HET INTERESSEBEGRIIP

1.1. *Begripsbepaling*

Het begrip interesse is één van de vele, waarop de uitspraak van toepassing is, dat zij in het dagelijks spraakgebruik nauwelijks toelichting behoeven, maar dat zij toch talrijke problemen blijken op te leveren, zodra men naar een goede definitie zoekt.

Deze problemen vinden we weerspiegeld in de ontwikkelingsgang, die het begrip in de psychologische literatuur heeft doorgemaakt.

Aanvankelijk heeft men geen zeer scherpe onderscheiding gemaakt tussen belangstelling en spontane opmerkzaamheid. Nog bij WM JAMES ⁽⁶⁹⁾ is de aandacht het belangrijkste criterium voor interesse. In de op hem volgende generatie vinden we, onder anderen bij DEWEY, echter reeds het inzicht, dat het interessebegrip een dubbel aspect heeft. Enerzijds is het een psychisch verschijnsel, waarin vooral de gevoelscomponenten opvallen. Anderzijds duidt het op een handelingsrelatie met de buitenwereld. Door DEWEY wordt dit, meer compact dan duidelijk, samengevat in de definitie van interesse als „the accompaniment of the identification, through action, of the self with some object or idea, because of the necessity, of that object or idea for the maintenance of a self-initiated activity” ⁽⁴⁴⁾, geciteerd bij BERDIE ²⁰⁾.

De behavioristische wending, die de Amerikaanse psychologie vervolgens heeft doorgemaakt, heeft zijn invloed ook op de opvattingen over het interessebegrip doen gelden. Weliswaar leent dit begrip zich niet zeer gemakkelijk voor een klassiek-behavioristische benadering en men heeft de emotionele aspecten er van ook nooit geheel uit het oog verloren. Maar het accent in de beschouwingswijze werd toch verlegd van de innerlijke componenten naar het uitwendig aspect: de belangstellingsstimuli, d.w.z. de personen, objecten, ideeën of situaties, die de belangstelling activeren, en naar de actie, die zij eventueel uitlokken. FRYER, vermoedelijk wel de eerste auteur, die een werk geheel aan de belangstellingspsychologie heeft gewijd, gaf in 1931 de definitie: „Subjective interests and aversions . . . are defined in terms of objects and activities stimulating the individual to have feeling experiences, and this experience can be estimated by the person being stimulated as to its quality of pleasantness or unpleasantness, and as to the intensity or degree of feeling” ⁽⁴⁹⁾. Deze definitie is

later in Amerika nogal eens met instemming geciteerd. We zullen hier op de merites van de omschrijving niet verder ingaan; alleen wijzen we er op, dat het laatste gedeelte van de begripsbepaling speciaal van belang is voor degenen, die de interessen experimenteel willen onderzoeken. Er wordt immers gezegd, dat de persoon in staat is tot een kwalitatieve en kwantitatieve schatting van de begeleidende emoties en impliciet dus tot ordening van zijn subjectieve interessen op grond daarvan.

Hoewel in Europa het belangstellingsbegrip met name in de pedagogische psychologie wel een belangrijke rol heeft gespeeld, men denke b.v. aan de „centres d'intérêt” van DECROLY, zijn vele theoretische psychologen toch aan het belangstellingsbegrip voorbijgegaan. Een uitzondering onder de oudere psychologen is W. STERN⁽¹¹⁶⁾, die in zijn uiteenzetting over de disposities, die een duurzame richting geven aan het menselijk voelen en daarmee ook aan het streven en handelen, onderscheid maakt tussen twee hoofdvormen: neigingen en interessen. Hij omschrijft neiging als „ein 'Haben' im Gefühl”, tegenover interesse als „ein 'Suchen' mit Gefühlsbeteiligung”. Hij voegt hieraan toe, dat interessen een grotere mate van distantie tot het object hebben, dat zij koeler en rationeler zijn dan de neigingen en daardoor ook meer probleembewustzijn geven. Deze toevoegingen, en ook de voorbeelden, die STERN daarbij geeft, maken wel duidelijk, dat de door hem gemaakte begripsonderscheiding, hoewel in de grond misschien wel juist, toch voor ons doel veel te subtiel is. Allerlei fenomenen, die men volgens de opvattingen van STERN tot de neigingen zou moeten rekenen, worden niet alleen in de gangbare opvatting, maar ook in alle onderzoeksmethodieken steeds als interessen beschouwd. We noemen hier als voorbeeld de zeer weinig probleembewuste, weinig rationele of gedistancieerde toewending van de vakman tot zijn arbeid, die we in het interessenonderzoek toch zeker als manifestatie van beroepsinteresse zullen opvatten.

In de na-oorlogse Europese opvattingen vinden we een poging, meer aansluiting te krijgen bij het Amerikaanse denken. Dit geschiedt door gebruik te maken van het begrip *instelling* („Einstellung”, „attitude”), waarvan de nauwe relatie met het interessebegrip onder anderen door IRLE⁽⁶⁸⁾ en — wat minder expliciet — ook door HOFSTÄTTER⁽⁶³⁾ is onderkend. Dezelfde conceptie is in Amerika door BERDIE⁽⁶⁰⁾, pag. 305 e.v.) naar voren gebracht in zijn omschrijving: „The term (interest) is used to designate a concept pertaining to factors within the individual which attract him to or repel him from various objects, persons and activities in his environment. The concept is primarily an affective one”. Dit is geheel in overeenstemming met de definitie van „attitude” als: „an enduring organization of motivational, emotional, perceptual and cognitive processes with respect to some aspect of the individual's world”⁽⁷⁸⁾. BERDIE onderscheidt dus niet alleen positieve interessen, maar ook negatieve (aversies).

In de hiervoor weergegeven opvattingen, die het interessebegrip op telkens andere wijze benaderen, zijn geleidelijk een aantal belangrijke kenmerken van dit begrip naar voren gekomen. Het zwaarste weegt o.i. wel de gerichtheid van het individu op bepaalde categorieën situaties, met uitlokkingskarakter, in de buitenwereld. Deze gerichtheid gaat uit boven de concrete, momentane beleving, zij draagt een min of meer permanent karakter. De beleving, die samengaat met deze gerichtheid en de daaruit eventueel voortvloeiende innerlijke of uitwendige actie, heeft een emotioneel aspect. In het algemeen denken we bij interessen vooral aan die belevingen, welke betrekkelijk dicht bij de oppervlakte van het psychisch leven liggen, zodat zij gemakkelijk bewust kunnen worden gemaakt.

Door deze laatste toevoeging wijzen we er op, dat de interessen deel uitmaken van een groter psychisch geheel, dat we kunnen aanduiden als de motivatie. Hieronder vallen, behalve de interessen ook allerlei onbewuste en de instinctieve drijfveren. Houden we hiermee rekening, dan komen we tot de volgende definitie:

De interessen vormen een aspect van de motivatie, dat bestaat uit de bewust-beleefbare, (semi-) permanente gerichtheid op bepaalde categorieën emotioneel stimulerende en tot actie uitlokkende situaties.

Deze definitie is van algemeen psychologische aard. Zij laat zich in de praktijk van het psychologisch onderzoek niet gemakkelijk toepassen. Wel kan echter dank zij deze begripsbepaling een operationele definitie worden verkregen, die practisch bruikbaar is. Hierop zal evenwel eerst later (paragraaf 7.2) worden ingegaan.

1.2. Karakteristiek der interessen

Vanzelfsprekend is het geven van een definitie niet voldoende om alles te zeggen, wat er over de achtergrond van het begrip interesse kan worden opgemerkt. Een analyse van de samenhang van dit begrip met andere psychische structuren, met name ook met de dieptepsychologische aspecten van de persoonlijkheid, valt echter buiten het bestek van deze studie. Datzelfde geldt ook voor de fenomenologie van de zich-interesserende mens. We willen hier uitsluitend aandacht schenken aan enkele aspecten, die in het bijzonder voor de belangstellingskarakteristiek van de jeugdige mens, omstreeks de overgang van lager naar voortgezet onderwijs, van betekenis zijn.

De concrete interessen ontstaan in de loop van de ontwikkelingsgang van het individu, door differentiatie van de persoonlijkheid. IRLE⁽⁹⁸⁾ heeft er op gewezen, dat in dit verband de opvattingen van LEWIN over de persoonlijkheidsontwikkeling⁽⁹⁹⁾ een vruchtbaar inzicht geven. Uitgaande van een „neutrale”, d.w.z. niet-individueel karakteristieke situatie, treedt geleidelijk een specificering van de relaties met de omgeving op. Deze krijgen onder anderen de vorm van interessen. In de periode van jeugd naar volwassenheid nemen deze in aantal geleidelijk toe, ondanks het feit, dat

ook bepaalde interessen bij het ouder worden verdwijnen. Er is echter niet alleen sprake van een kwantitatieve vermeerdering, maar ook van een voortschrijdende innerlijke organisatie van de interessen. Met de leeftijd neemt de complexiteit van de interessen toe, er ontstaan o.a. binnen de grote interessegebieden specifieke deelinteressen. Bovendien worden de interessen steeds meer geordend in een hiërarchie, waarbij bepaalde interessen aan andere dienstbaar worden gemaakt. Ook de onderlinge vervlochtenheid van de interessen, het gericht zijn op verwante, elkaar eventueel overlappende situaties, neemt toe. Kenmerkend voor de ontwikkeling is voorts de vergroting van de „life space”: het aantal sectoren waarvoor men zich zou kunnen interesseren wordt groter; van de „space of free movement”: het aantal situaties, waarin men zou kunnen ageren neemt toe; en van de „time perspective”: de permanentie van de gerichtheid komt steeds duidelijker tot uiting. Tenslotte blijkt de ontwikkeling ook uit een toenemend realisme van de interessen. Deze kenmerken, die volgens LEWIN algemene kenmerken van de persoonlijkheidsontwikkeling zijn, gelden dus ook voor de specifieke ontwikkeling van de interessen.

Het eerste ontstaan van de interessen blijft in deze theorie onduidelijk. Uit een onderzoek van TYLER (1933), dat in het bijzonder op de genese van de beroepeninteressen betrekking had, is nu de opvatting naar voren gekomen, dat het zeer jeugdige kind ongedifferentieerd positief staat ten opzichte van de veelheid van mogelijkheden. Een belangstellingsstructuur ontstaat hieruit in eerste aanleg, doordat er een afwending van bepaalde mogelijkheden optreedt. Eerst daarna zien we de positieve toewending tot andere mogelijkheden geaccentueerd worden. Dit proces is, blijkens de onderzoekingen van TYLER, voorzover het de beroepeninteressen betreft, op tienjarige leeftijd reeds in volle gang.

Over de ontwikkeling van de belangstelling in positieve zin is vooral door CARTER (28, 29) in Amerika studie verricht. Ook hij heeft zich speciaal gericht op de interessen, die verband houden met de voorbereiding op de beroepskeuze. Gebleken is, dat de positieve ontwikkeling van deze interessen, vooral onder invloed van de identificatie met bepaalde personen en groepen tot stand komt. Onder deze invloed gaat men zich speciaal op bepaalde gebieden oriënteren, krijgt men een voorkeur voor bepaalde typen ervaringen. Eventueel gaat hiervan zelfs enige modificerende werking op de persoonlijkheid uit. Volgens CARTER kan de leeftijd waarop de interessen overwegend een vorm hebben gekregen, die past bij de totale psychische, lichamelijke en sociale situatie van de persoon, echter individueel sterk uiteenlopen. Hier kan in het bijzonder de opvoedende werking van het milieu een belangrijke rol spelen.

In de loop van de ontwikkeling verandert ook de persoonlijke betekenis van de interessen. Zij kunnen functioneel autonoom worden, in de zin, waarin ALLPORT deze term gebruikt. Uiteindelijk kunnen zij zelfs tendele een „objectief” karakter krijgen, zoals de „levenshoudingen” van SPRANGER

(113), die immers ook als interessen kunnen worden opgevat. Bij de groep, die ons hier in het bijzonder interesseert, is dit laatste ontwikkelingsstadium nog niet bereikt, de interessen hebben hier nog een uitsluitend subjectief-functioneel karakter.

Ten aanzien van de inhoudelijke kant van de belangstellingskarakteristiek rijzen ook vragen. Wijkt de geaardheid van de belangstelling van de 12- tot 14-jarigen belangrijk af van die van andere, in het bijzonder oudere leeftijdsgroepen? Dat er een leeftijdsontwikkeling in de belangstellingskarakteristiek bestaat, is macroscopisch onmiddellijk waar te nemen, bij vergelijking van sterk uiteenlopende leeftijdsgroepen. Interessen verschijnen en verdwijnen, en ook als zij naar de vorm blijven bestaan, richten zij zich toch op andere objecten (b.v. in het spel, de lectuur, e.d.).

Toch is het niet gemakkelijk, deze macroscopisch waarneembare ontwikkeling voor bepaalde leeftijdsgroepen te preciseren. De voornaamste moeilijkheid blijkt wel te zijn, dat een zinvolle rubricering van de interessen complicaties oplevert. Het aantal mogelijke belangstellingsvormen is in de praktijk vrijwel onbegrensd en de ordening, die men in deze veelheid zou willen aanbrengen, is sterk subjectief gekleurd. Dit bemoeilijkt de unificatie van het grote aantal onderzoeken, dat toch nodig zou zijn, om de leeftijds-karakteristiek van een aanzienlijk aantal groepen goed te leren kennen.

Het is daarom niet verwonderlijk, dat onze kennis van de specifieke interessen van de 12- tot 14-jarigen nog slechts zeer oppervlakkig is, zelfs als we ons beperken tot de interessen, die van directe betekenis zijn voor de school- of beroepskeuze.

Een moeilijkheid rijst ook in verband met de wijze, waarop de interessen aan de onderzoeker kenbaar worden. SUPER⁽¹²⁷⁾ heeft een nuttige onderscheiding gemaakt in „manifest”, „expressed” en „inventoried” interessen, drie vormen waarvan in de praktijk is gebleken, dat zij onderling slecht correleren. Onder de „manifest interests” verstaat hij de belangstellingsvormen, waarvan het bestaan door directe gedragsobservatie kan worden vastgesteld, dus b.v. door waarneming van spel, liefhebberijen, tijdsbesteding, e.d. Deze interessen schijnen echter noch met de beide andere vormen, noch met de latere beroepskeuze veel verband te houden, behalve in het geval van enkele zeer krachtige, reeds vroeg permanent-wordende interessen. In onze studie zullen zij verder buiten beschouwing worden gelaten.

Onder „expressed interests” worden de interessen verstaan, die door het betrokken individu verbaal kenbaar worden gemaakt, bij ondervraging. Men weet hiervan reeds geruime tijd, dat zij betrekkelijk weinig betrouwbaar zijn⁽⁸⁸⁾. Bij vele kinderen is de rechtstreekse mededeling over de belangstelling sterk afhankelijk van momentane invloeden, zodat bij herhaling van het onderzoek reeds na betrekkelijk korte tijd aanzienlijke veranderingen in de onderzoekresultaten voorkomen. Ook in dit geval is

de verwantschap tussen de kenbaar gemaakte beroepsinteressen en de latere beroepskeuze van geringe praktische betekenis. Toch zijn er juist over deze vorm van interesses verscheidene onderzoeken verricht. Hiertoe behoren een aantal oudere Duitse studies, die b.v. betrekking hebben op de voorkeur voor schoolvakken. Van een aantal van deze heeft LANGEVELD ⁽⁸⁴⁾ een overzicht gegeven. Voor ons doel hebben zij echter weinig betekenis, omdat de groepen waarop het onderzoek betrekking heeft gehad gewoonlijk te vaag omschreven zijn, zodat de representativiteit van de bevindingen dubieus is.

Wat meer waarde kan worden gehecht aan de onderzoeken over de concrete beroepskeuzewensen van de kinderen, die de lagere school gaan verlaten. Daarbij worden deze wensen als uitingen van interesses opgevat. Sinds de klassieke studies van de Weense groep ⁽⁸⁶⁾ zijn er verschillende onderzoeken over dit onderwerp geweest. In ons land zijn zij, met betrekking tot leerlingen van lagere technische scholen, verricht door BORDEWIJK ⁽²²⁾. In overeenstemming met andere onderzoekers, in Duitsland, Oostenrijk en Schotland, vond hij bij zijn proefpersonen een overwegende voorkeur voor beroepen uit de metaaltechnische sector. Deze voorkeur is van betrekkelijk recente datum: nog tegen het einde van de vorige eeuw ging, althans in ons land, de voorkeur zeer overwegend uit naar de traditionele ambachtelijke beroepen, waaronder vooral de houtbewerking in de gunst stond. De verschuiving van de preferentie kan statistisch zeer fraai worden gedemonstreerd.

De bedoelde onderzoeken hadden betrekking op een speciale categorie kinderen, die kan worden omschreven als het deel van de grotestadsjeugd, dat een handarbeidersberoep kiest, zodat de uitkomsten dus niet zonder meer gegeneraliseerd mogen worden. Toch kan op grond van observaties wel het vermoeden worden uitgesproken, dat omstreeks het einde van de schoolkindperiode en in de vroege puberteit de toewending tot praktisch-zakelijke situaties, althans bij jongens, overheerst. De nadruk valt enerzijds op belangstellingsstimuli, die een doe-element bevatten en verder op de prikkels tot rationeel weten. Daartegenover staat een afwending van situaties met geringe uitwendige actie-mogelijkheden (kantoorwerk!) en van emotioneel-belastende situaties. Dit laatste geldt voor de meisjes zeker niet, in tegendeel prevaleren bij hen duidelijk de interesses, die verband houden met sociale en pedagogische problemen (de bekende voorkeur voor beroepen als verpleegster, kleuterleidster, kinderverzorgster, e.d.). Op grond van observaties en interviews is ook wel waarschijnlijk, dat er bij de jongens een ontwikkeling is, die inhoudt, dat omstreeks het midden van de puberteit ook de geesteswetenschappelijk of esthetisch gerichte belangstelling wat meer tot uiting komt en dat nog later, in de adolescentie, ook het sociaal interesse ontwaakt. STRONG ⁽¹²⁴⁾, die in het kader van de door hem geconstrueerde belangstellingstest ook een „maturity-scale” heeft gemaakt, heeft daarin vooral items opgenomen,

die sterk blijken te correleren met belangstelling voor maatschappelijk werk. Experimenteel is gevonden, dat bij jongens de belangstelling voor dergelijke items in de loop van het laatste deel van de puberteit significant toeneemt ⁽¹¹⁰⁾. Bij meisjes, bij wie de gemiddelde belangstelling voor deze items reeds vroeg zeer groot is, valt een dergelijke ontwikkeling niet waar te nemen. We kunnen hier op grond van eigen onderzoek nog aan toevoegen, dat wij de indruk hebben, dat deze ontwikkeling bij de jongens niet rechtlijnig verloopt, in de zin van een geleidelijk toenemende belangstelling voor sociale en pedagogische situaties. Eerder moet worden aangenomen, dat in de eerste helft van de puberteit nog een verdergaande afwending optreedt, die pas later in het tegendeel verkeert.

Wat tenslotte de „inventoried interests” betreft, hiertoe worden gerekend de interesses, waarvan het bestaan met behulp van een test, in de meeste gevallen een vragenlijstmethode, is vastgesteld. Hiervan is gebleken, dat zij wel relevant zijn in verband met de beroepskeuze ^(124, 125). Bovendien hebben zij andere voordelen. De betrouwbaarheid van de uitkomsten is groter, omdat tests minder kwetsbaar blijken voor de invloed van momentane situaties, dan de uitspraken, die in een vraaggesprek worden gedaan. Verder maken vele tests gebruik van vaste belangstellingscategorieën, die bovendien bij verschillende tests niet zeer sterk uiteenlopen. Daardoor wordt ook eerder de mogelijkheid geboden, een systematisch onderzoek over de belangstellingskarakteristiek van een bepaalde leeftijdsgroep te doen. Toch is er ook wat de „inventoried interests” betreft, nog niet al te veel bekend over de groep, waarvoor wij ons in het bijzonder interesseren. In Amerika wordt zij als regel nog niet in het belangstellingsonderzoek betrokken, omdat daar de overgang van het lager naar het voortgezet onderwijs niet met een differentiatie naar schooltype gepaard gaat. In Europa is daarentegen het belangstellingsonderzoek eerst betrekkelijk laat op gang gekomen. Hier heeft men wel aandacht aan de 12- tot 14-jarigen geschonken, maar het gepubliceerde materiaal is nog te fragmentarisch, om een samenhangend beeld te kunnen vormen. Uit ons eigen onderzoek, dat in de tweede helft van dit boek zal worden gerapporteerd, komen echter gezichtspunten naar voren, die het mogelijk maken een beter inzicht in de belangstellingskarakteristiek van deze leeftijdsgroep en ook van verschillende subgroepen binnen deze categorie te krijgen.

1.3. *Stabiliteit der interesses*

De laatste vraag, waarmee wij ons in dit hoofdstuk zullen bezighouden, betreft de waarde op langere termijn van de resultaten, die bij interessesonderzoek worden verkregen. Aanvankelijk heeft men vooral op grond van onderzoek dat betrekking had op „expressed interests” gemeend, dat de stabiliteit der belangstelling zo gering was, dat het vrijwel uitgesloten zou zijn, de belangstelling als een individuele predictor, b.v. in het kader van voorlichting bij school- en beroepskeuze, te gebruiken. Op deze pessi-

mistische opvatting is men later, toen men zich meer op de „inventoried interests” ging richten, wel enigszins teruggekomen. Maar toch rijzen hier problemen. Deze problemen kunnen ook op grond van een aprioristische redenering wel voorzien worden. Onderzoekingen in verschillende landen (43, 51, 86) hebben immers aangetoond, dat er in de verlangens met betrekking tot de beroepskeuze eerst gemiddeld omstreeks het 16e à 18e levensjaar een aanzienlijke mate van stabiliteit begint te komen. Tot deze leeftijd maakt men een ontwikkelingsproces door, waarin men geleidelijk voor allerlei aspecten van de arbeid en van het beroep begrip begint te krijgen. In samenhang hiermee kan de voorkeur voor het te kiezen beroep telkens veranderingen ondergaan. Het ligt nu wel voor de hand, dat dit ontwikkelingsproces ook zijn invloed zal laten gelden op de individuele ontwikkeling van de belangstellingskarakteristiek. Daargelaten of er nu grote leeftijdsgebonden verschuivingen in de belangstellingskarakteristiek voorkomen of niet, zullen er in ieder geval persoonlijke factoren zijn, die kunnen meebrengen, dat de voorkeur, zoals deze in een belangstellingstest tot uitdrukking komt, niet constant is.

Over de stabiliteit van de uitkomsten van belangstellingstests is in de Verenigde Staten vrij veel onderzoek verricht. De zeer uitvoerige studies van STRONG (124, 125) hebben wel duidelijk gemaakt, dat er bij de jonge volwassenen, in het bijzonder de *college*-studenten, een fase is bereikt, die individuele predictie mogelijk maakt. De uitkomsten bij follow-up na zeer aanzienlijke tijd (10 à 20 jaar na het oorspronkelijke onderzoek) leerden, dat verschillende criteria inderdaad een zeer bevredigende overeenstemming met de uitkomsten van de belangstellingstest vertoonden.

Bij jeugdiger personen is de overeenstemming echter minder bevredigend. Op grond van een literatuuronderzoek kwam ARCHER (100, pag. 1162 e.v.) tot de gevolgtrekking, dat er over een tijdsverloop van één jaar bij de jongens 60% en bij de meisjes 70% van de interessen constant blijft. De leeftijd, waarop deze bevindingen betrekking hebben, wordt niet meegedeeld; op basis van enkele der door hem geciteerde publicaties mag echter wel worden aangenomen, dat het gemiddeld ongeveer 16-jarigen betrof. STORDAHL (geciteerd bij 41, pag. 39 e.v.) vond bij hertesten na twee jaar van een groep, die op ongeveer 17-jarige leeftijd was onderzocht, een gemiddelde correlatie op de schalen van de STRONG Vocational Interest Blank van + .70. In overeenstemming hiermee vond TAYLOR (128) bij hertesten na twee jaar een correlatie op dezelfde test van gemiddeld .73 bij oorspronkelijk 18-jarige, maar daarentegen van gemiddeld .60 bij oorspronkelijk 16-jarige jongens. ROSENBERG (110) vond zowel bij jongens als bij meisjes een mediale correlatie van .61 op negen schalen van de KUDER Preference Record, bij hertesten na drie jaar. Het eerste onderzoek had in de negende klas, dus op ongeveer 15-jarige leeftijd plaatsgevonden. Over nog jeugdiger leeftijdsgroepen zijn weinig gegevens bekend. Vermeld wordt een niet gepubliceerde studie van CARTER (geciteerd bij 41, pag. 38),

betrekking hebbende op 11-jarige jongens. In hun geval was de correlatie bij hertesten na één jaar .62, na twee jaar .51 en na drie jaar .31.

Deze bevindingen zijn onderling consistent. Zij wijzen er op, dat bij het ouder worden de stabiliteit van de voorkeur geleidelijk groter wordt en dat vooral de stabiliteit op langere termijn aanzienlijk verbetert. Ter vergelijking met de genoemde waarden geven we enkele door STRONG ⁽¹²⁴⁾ gevonden correlaties bij heronderzoek van personen, die aanvankelijk als vierdejaars *college*-studenten waren onderzocht. Na vijf jaar was de gemiddelde correlatie nog .75, na tien jaar .71. De veel grotere constantheid op langere termijn van deze oudere groep wordt daarmee duidelijk geïllustreerd.

Ten aanzien van de door ons bestudeerde leeftijdsgroep moeten we, op grond van deze Amerikaanse onderzoeken, geen al te hoge verwachtingen hebben over de stabiliteit van de scores op langere termijn. Weliswaar zijn er, zoals onder anderen een onderzoek van TYLER ⁽¹³²⁾ heeft aangetoond, voorkeursverschillen tussen Amerikaanse en Europese (Engelse) kinderen. De laatsten bleken een kleiner aantal sympathieën en een groter aantal antipathieën te hebben, dan de eerstgenoemden. In het licht van TYLER's eerder vermelde opvattingen over de genese van de belangstellingsstructuur, zouden wij dit kunnen opvatten als een aanwijzing dat bij de onderzochte Europese kinderen deze structuur in een verder gevorderde ontwikkelingsfase verkeerde. Maar het is de vraag, of hieruit zou mogen worden geconcludeerd, dat bij Europese kinderen veel vroeger een hoge graad van stabiliteit in het belangstellingspatroon optreedt, dan in Amerika het geval is.

Wel is er een andere overweging, die daarvoor tot op zekere hoogte pleit: de differentiatie, die na de lagere school wordt doorgevoerd, heeft ongetwijfeld een fixerende werking op de belangstelling. Zoals later zal blijken vinden we in de belangstellingstest reeds bij de leerlingen van de zesde klasse verschillen, die duidelijk correleren met de vorm van voortgezet onderwijs, die zij gaan kiezen. De kandidaten voor het v.h.m.o. hebben een meer dan gemiddelde belangstelling voor de „intellectuele” items uit de test, de landbouwers in spe voor de agrarische items, de aspirant-metaalbewerkers voor items, die betrekking hebben op technische handarbeid. In ieder van deze gevallen schept de omgeving, waarin zij na de lagere school binnentreden een milieu, waarin juist de genoemde activiteiten en belangstellingsstimuli een centrale plaats hebben. Er is daardoor a priori een aanzienlijke kans, dat de leerlingen, die zich in het bijzonder met deze onderwerpen bezighouden, aan de betreffende items ook bij heronderzoek een aanzienlijke voorkeur geven.

Welke kwantitatieve betekenis echter aan het effect van een dergelijke fixerende werking moet worden toegekend, valt zonder nader onderzoek niet te beoordelen. Het zou daarom zeker wenselijk zijn, indien ook in ons land eens een onderzoek naar de constantheid van de belangstelling, in

het bijzonder bij de leeftijdsgroep, waarvan het onderzoek in de praktijk een grote rol speelt, zou geschieden.

We moeten er tenslotte op wijzen, dat zelfs indien een dergelijk onderzoek zou aantonen, dat de interessen van jeugdigen op langere termijn slechts een geringe stabiliteit zouden hebben, daarmee toch nog niet de waarde van het psychodiagnostisch belangstellingsonderzoek met het oog op de predictie van school- of beroepssucces wordt aangetast. Dit is immers primair een kwestie van predictieve validiteit en deze moet afzonderlijk onderzocht worden. Nu bestaat er weliswaar tussen de stabiliteit, analoog aan de betrouwbaarheid, en de validiteit van een test wel een theoretische relatie, maar deze is van dien aard, dat ook bij een relatief bescheiden stabiliteit toch nog een aanzienlijke predictieve validiteit mogelijk is. Wel zal de psycholoog er verstandig aan doen, zich bij jeugdigen te onthouden van voorspellingen op langere termijn. Dit geldt echter niet alleen in het kader van het gebruik van belangstellingstests, maar evenzeer bij toepassing van vrijwel alle andere psychologische hulpmiddelen. Bij predicties op korte termijn, b.v. over de ontwikkeling in de eerstvolgende één à twee jaar, is de toepassing van belangstellingstests ook bij een relatief bescheiden mate van constantheid der interessen zeker niet bij voorbaat af te wijzen. Op een minder exacte wijze, die echter beter aansluit bij de gangbare denkwijze, kan men hetzelfde zeggen, door te stellen, dat zelfs als de interessen weinig constant zijn, het toch aanbeveling kan verdienen, bij de voorlichting bij school- en beroepskeuze mede rekening te houden met de motiverende krachten, die actueel bij het kind aanwezig zijn.

HOOFDSTUK II

INTERESSENONDERZOEK

2.1. *Historisch overzicht*

De geschiedenis van het belangstellingsonderzoek is reeds een aantal malen op uitstekende wijze beschreven (²⁰, ⁴¹, ⁴⁹, ¹³¹). Daarom zal hier worden volstaan met enige, in ons land over het geheel weinig bekende hoofdpunten te releveren.

De geboekstaafde geschiedenis van het belangstellingsonderzoek begint dan kort na 1910 in de Verenigde Staten, waar KELLEY, destijds verbonden aan het *Teachers College* van de Columbia Universiteit, een eerste ontwerp van een belangstellingstest publiceerde (⁷²). Deze test, die gezien vanuit de tegenwoordige stand van de techniek zeer primitief van constructie was, is toch belangwekkend, omdat zij reeds een aantal grondprincipes bevat, die ook later bij de ontwikkeling van methoden voor belangstellingsonderzoek een rol hebben gespeeld. KELLEY's test begint met een lijst van tijdschriften; de proefpersonen — de test was bedoeld voor leerlingen van het voortgezet onderwijs — moeten eerst aangeven, welke periodieken hen bekend zijn en vervolgens, in volgorde van voorkeur, de vijf, die zij de interessantste vinden. Dan volgen vragen naar de drie interessantste boeken, die zij in de laatste twee jaren hebben gelezen, de meest geliefde binnen- en buitensport, de meest aantrekkelijke uit 14 met name genoemde vormen van ontspanning en de twee beroepen waarnaar de voorkeur het meest uitgaat. De test wordt besloten met een woordenschat-test.

Dit werk van KELLEY schijnt, zelfs binnen de Verenigde Staten, betrekkelijk weinig aandacht te hebben getrokken. Ook de meeste tests, die in de jaren na de eerste wereldoorlog werden ontwikkeld, zijn, in enkele gevallen na een periode van betrekkelijk grote populariteit, weer in de vergetelheid verdwenen.

Vermelding verdient nog een „seminar”, dat onder leiding van YOAKUM, in het cursusjaar 1919/20 aan de Columbia Universiteit werd gehouden. Een van de doelstellingen van dit „seminar” was het verzamelen van testvragen voor het belangstellingsonderzoek bij verschillende leeftijdsgroepen. Het omvangrijke materiaal, in totaal werden ca. 3.000 testvragen voor uiteenlopende leeftijden verzameld, is de basis geweest voor de constructie van verschillende tests. Hiertoe behoorde o.a. de *Carnegie Interest Inventory*, een vragenlijst die bestond uit 72 beroepen en 126 andere items, betrekking hebbende op eigenschappen van mensen. Iedere item moest

worden beoordeeld op een driedelige schaal: „Like”, „Indifferent”, „Dislike”. De keuze op deze schaal wordt in de literatuur dikwijls aangeduid als de LID-techniek; deze speelt ook thans in een aantal belangstellingstests nog een rol.

De Carnegie-test is op zijn beurt het uitgangspunt geweest voor het werk van een aantal psychologen, verbonden aan de Stanford Universiteit, aanvankelijk vooral onder leiding van FREYD en COWDERY, later onder die van STRONG. In de loop van enige jaren verschenen uit hun kring een aantal nieuwe bewerkingen van de test, die vooral werden gekenmerkt door het feit, dat het aantal items groter was en de verscheidenheid van de gebieden, waaruit de items werden gekozen, eveneens groter. De LID-techniek bleef gehandhaafd.

Een voorlopig eindpunt van de ontwikkeling werd in 1928 bereikt met de verschijning van het eerste model van de *Strong Vocational Interest Blank*, die sedert de voornaamste plaats onder de Amerikaanse belangstellingstests heeft veroverd. In zeer vele opzichten komt dit eerste model nog goed overeen met de — later te bespreken — hedendaagse uitvoering. De grote verdienste van STRONG is vooral de zorg, die hij aan de statistische bewerking van zijn test en in het bijzonder aan de validatie besteedde. Vooral in dit opzicht onderscheidde hij zich van zijn tijdgenoten, die dikwijls toch wel tests met verdienstelijke kwaliteiten hebben geproduceerd. Het schijnt niet te gewaagd, te veronderstellen, dat in dit verschilpunt de reden ligt, waarom de test van STRONG zich, als enige uit de begintijd, heeft weten te handhaven.

In de periode na 1930 treedt er een wat grotere divergentie in de ideeën op, die ten grondslag worden gelegd aan de nooit aflatende stroom van nieuwe interessentests. De meeste van hen hebben nooit een aanzienlijke praktische betekenis verkregen. Een vergelijking b.v. van de tweede en de vierde uitgave van het bekende handboek van BUROS^(25, 27) leert, dat er ten aanzien van de interessentests slechts een zeer geringe overlapping bestaat. In een periode van minder dan vijftien jaar hebben slechts drie tests zich weten te handhaven. Uit de tijdschriftliteratuur blijkt bovendien, dat ook de tests van recente datum meest eendagsvliegen zijn: er wordt over toepassing van deze tests nimmer gerapporteerd.

Hoofdvorm van de latere verschijningen op het terrein van de belangstellingstests is nog steeds de vragenlijst. Een nieuwe uitvoering hiervan is echter de onderlinge vergelijking van een aantal items op hun aantrekkelijkheid. Indien deze items ontleend zijn aan verschillende interessegebieden, bestaat de mogelijkheid, zodoende de relatieve sterkte van de affiniteit tot deze gebieden te bepalen. De belangrijkste toepassing van dit principe vinden we in de in 1939 verschenen KUDER *Preference Record*, die zich naast de test van STRONG een erkende positie in het psychodiagnostisch arsenaal heeft weten te verwerven. Ook deze test zal in de volgende paragraaf worden besproken.

Naast de vragenlijstvorm zijn sedert ook enkele andere testvormen gebruikt voor het belangstellingsonderzoek. Reeds van betrekkelijk oude datum (1925) is de poging, een woordassociatietest te gebruiken voor dit doel; het is echter tot dusver niet gelukt, een praktisch hanteerbare vorm voor de beoordeling van de reacties op de prikkelwoorden te vinden (⁴⁹, ¹⁴²). GREENE (⁵³) heeft de oude gedachte van KELLEY, een woordenschattest te gebruiken om uit de niveauverschillen in de bekendheid met verschillende woorden conclusies te trekken over de belangstellingsstructuur, nader uitgewerkt. Maar ook deze methode heeft geen vaste voet gekregen in de onderzoekkamer van de psycholoog. Hetzelfde geldt voor een poging, de belangstellingskarakteristiek met behulp van een projectieve methode, gebaseerd op het principe van de *Thematic Apperception Test* te benaderen (¹, ²). Wel zou men echter, op grond van het werk van McCLELLAND en zijn medewerkers (⁹⁵), de verwachting kunnen uitspreken, dat het onderzoek van de fantasie, met behulp van projectieve methoden, in de toekomst een grotere bijdrage zal leveren tot de praktische kennis van de motivatie.

De hiervoor aangeduide ontwikkeling heeft zich geheel in de Verenigde Staten afgespeeld. Daarbuiten is door de psychologen lange tijd vrijwel geen aandacht aan dit gebied besteed. In Europa werd als enig hulpmiddel bij het interessenonderzoek de beroepenlijst gebruikt, waarop volgens het principe van de vrije keuze een voorkeur tot uitdrukking kon worden gebracht. In Nederland zijn dergelijke beroepenlijsten nog steeds zeer populair, om deze reden zal er in de volgende paragraaf nader op worden ingegaan. Vooral in Zwitserland zijn voorts vele varianten op het principe van de beroepenlijst bedacht, waarbij in plaats van beroepen b.v. materialen of bezigheden ter vrije keuze werden gegeven. Nauw verwant aan dit principe is ook dat van de „boekencatalogustest” van BAUMGARTEN (¹⁶, ¹⁷), die in de Duits sprekende landen nog wel enige toepassing schijnt te vinden. Kenmerkend voor al deze pogingen is, dat zij van testmethodisch gezichtspunt gezien als zeer primitief moeten gelden. Onder anderen ontbreekt iedere poging tot statistische toetsing van de testprincipes en -uitvoering.

Eerst enkele jaren na de tweede wereldoorlog begon in Europa de ontwikkeling van belangstellingstests van beter kaliber. Bij een internationale uitwisseling van ervaringen (⁴) is kort geleden gebleken, dat thans in vele landen werk op dit gebied wordt verricht. Daarbij sluit men zich in het algemeen nauw aan bij het Amerikaanse voorbeeld. Ook Europa kent een poging tot constructie van een projectieve belangstellingstest (¹⁸), die vooral in Italië de aandacht schijnt te hebben getrokken. De overige onderzoekers maken echter gebruik van de vragenlijstmethode, waarbij het werk van STRONG of KUDER meestal het uitgangspunt vormt.

RENNES (¹⁰⁸) heeft in Frankrijk een test geconstrueerd, waarbij evenals bij de STRONG-test iedere item afzonderlijk wordt beoordeeld. Zijn werk

is vooral interessant door zijn poging, niet alleen de relatieve, maar ook de absolute sterkte van de interessen te bepalen; zij het ook, dat men, wat dit laatste betreft, een discussie van de vele netelige methodologische aspecten van een dergelijke poging node mist. Op het werk van RENNES is verder gebouwd door de Zwitser DUPONT (4), die vooral onderzoekingen over de validiteit van de test heeft verricht.

De KUDER-test, die in Frankrijk ook officieel vertaald is, heeft in dat land tevens als voorbeeld gediend voor de constructie van een belangstellingstest voor jongens van 13—16 jaar. CHAUDAGNE (37) en LARCEBEAU (35) hebben hierover gepubliceerd. Bovendien heeft LARCEBEAU een andere belangstellingstest voor jongens, ditmaal van 10—12 jaar ontworpen, die op het STRONG-principe is gebaseerd (4).

Min of meer op het KUDER-beginsel is voorts het werk van FITZPATRICK en WISEMAN in Engeland (48) gebaseerd. Zij combineren steeds in een serie van zes items één sociale, twee technische en twee intellectuele activiteiten met één „strong distractor” (b.v. voetballen). Uit deze serie moet een eerste en een tweede voorkeur worden aangegeven.

Ook in andere landen heeft de KUDER-test de aandacht getrokken. In België kwam een Vlaamse vertaling tot stand (39). In Duitsland heeft IRLE, wiens werk het uitgangspunt voor onze studie heeft gevormd, zijn test eveneens op het KUDER-principe gebaseerd.

Over de situatie in Nederland tenslotte is officieel weinig bekend. Naast de beroepenlijsten schijnen door enkele psychologen wel zelfvervaardigde vertalingen van buitenlandse tests, met name weer van de KUDER *Preference Record* te worden gebruikt. Voorts is bekend, dat bij het HOOGVELD-Instituut een belangstellingstest van eigen ontwerp werd geconstrueerd (12), nadere gegevens zijn hierover echter nog niet meege-deeld.

2.2. Hoofdvormen van tests

Uit het historisch overzicht blijkt, dat de vragenlijstmethode tot dusver in het belangstellingsonderzoek een zeer overheersende rol speelt. Om dit te begrijpen moeten we in de eerste plaats bedenken, dat met het begrip belangstelling een aspect van de bewuste motivatie wordt aangeduid. Het gaat om drijfveren en daarmee verbonden attitudes, die zo dicht bij de oppervlakte van het psychisch leven liggen, dat zij betrekkelijk gemakkelijk voor introspectie toegankelijk zijn. Ook de mens die betrekkelijk weinig afstand neemt tot zichzelf, die zijn reacties niet bewust analyseert en daardoor niet in staat is een helder oordeel over de bij hem bestaande interesseverhoudingen te geven, kan toch ten opzichte van concrete situaties zijn voorkeur of afkeer gemakkelijk bewust maken. In een dergelijke situatie is de vragenlijstmethode ook in andere gevallen, b.v. bij sociaal-wetenschappelijke onderzoekingen, een goed hulpmiddel gebleken. Daarbij heeft de vragenlijstmethode praktische voordelen. Zij leent zich, in

tegenstelling tot b.v. een projectieve methode, gemakkelijk tot klassikale toepassing. De uitwerking is, omdat bij ieder testonderdeel slechts uit enkele mogelijkheden hoeft te worden gekozen, gemakkelijk te standaardiseren, zodat de scoring machinaal of door inschakeling van hulpkrachten van eenvoudig niveau kan geschieden. Men kan de test ook bij intellectueel weinig geschoolde of zwakbegaafde personen afnemen, zonder veel kans op misverstanden. Al deze overwegingen steunen de conclusie, dat de vragenlijstmethode inderdaad een goed bruikbaar beginsel voor de constructie van een belangstellingstest vormt.

Ten aanzien van de concrete uitvoeringsvorm rijzen echter nog wel vragen. Deze zullen thans aan de orde worden gesteld in het kader van een bespreking van een drietal tests: beroepenlijsten, *STRONG-Vocational Interest Blank* en *KUDER-Preference Record*.

BEROEPENLIJSTEN

Beschrijving: De test bestaat uit een formulier, waarop een aantal beroepen staan vermeld. Dit aantal kan sterk variëren. Er zijn beroepenlijsten met 20 à 30 beroepen, maar ook die er vijf maal zoveel bevatten. De onderzochte wordt gevraagd, alle beroepen, die hem aantrekkelijk voorkomen, met een + teken aan te duiden en alle beroepen, die hem afstoten met een — teken. Veelal wordt bij de testinstructie gezegd, dat het financieel aspect van het beroep en de opleiding buiten beschouwing moet worden gelaten en dat ook niet moet worden gelet op de geschiktheid of de praktische mogelijkheid, een eventuele beroepswens te realiseren.

Kleine modificaties in de instructie komen voor. De onderzochte kan worden opgedragen, alle beroepen van een beoordeling +, 0 of — (L I D) te voorzien, of om alleen die beroepen, waarvoor een duidelijke voorkeur of afkeer aanwezig is, aan te geven. Eventueel kan ook worden toegestaan een zeer sterke sym- of antipathie met een groter aantal + of — tekens kenbaar te maken.

Een variant op dit principe is de beroepenlijst, waarin b.v. telkens drie beroepen van ongeveer gelijk maatschappelijk niveau naast elkaar zijn geplaatst, de onderzochte moet één hiervan kiezen (¹³⁶).

Hoewel zowel in binnen- als buitenland talloze modellen van beroepenlijsten in gebruik zijn, is hierover in de vakliteratuur nauwelijks een vermelding te vinden. Zeker zijn zij nooit het object van een diepergaande studie geweest. Wel zijn echter langzamerhand vele gebruikers van deze methode zich gaan realiseren, dat hieraan ernstige test-technische bezwaren kunnen zijn verbonden.

Deze bezwaren blijken reeds, indien men zich afvraagt, welke betekenis aan de voorkeursreacties moet worden gehecht. Men kan de beroepenlijst opvatten als een steekproef uit het veld van alle keuzemogelijkheden en in dit geval de voorkeursreactie opvatten als een *teken* (⁵²) voor een reële affiniteit tot het betreffende beroep. Men kan echter de test ook beschou-

wen als een projectieve methode en dan de voorkeursreactie nemen als een *symbol*, dat dient als uitgangspunt voor een psychologische interpretatie. Beide opvattingen vinden aanhang. Van dieptepsychologische zijde, met name uit de school van SZONDI ⁽¹⁰¹⁾ zijn pogingen gedaan, een symbolische interpretatie voor de beroepsvoorkeur te geven. Deze berusten echter op een systeem met zeer aanvechtbare grondslagen, waarin bovendien voortdurend wijzigingen worden aangebracht, zodat interpretaties na enige tijd hun verband met het systeem blijken te hebben verloren ⁽¹³⁵⁾. Dit zou misschien nog niet zo ernstig zijn, indien niet bovendien de relatie tussen de grondvariabelen van het SZONDI-systeem en de beroepsvoorkeur in vele gevallen niet doorzichtig en bovendien nooit experimenteel bewezen is. In hoofdzaak geldt hetzelfde voor pogingen tot symbolische interpretatie van freudiaanse zijde ⁽⁶⁴⁾, die weliswaar beter uitgewerkt zijn, maar toch nimmer gevalideerd.

We beperken ons daarom tot de opvatting van de voorkeursreactie als „teken”. Dit is in ons land ook algemeen gebruikelijk. De voorkeur, die de onderzochte uitsprekt — en m.m. geldt een analoge zienswijze voor de afkeer — wordt opgevat als een uiting van affiniteit tot de betreffende beroepsrealiteit. Als de onderzochte b.v. een +reactie geeft bij de beroepen van bakker en bankwerker, wordt dit geïnterpreteerd als een teken van sympathie voor de realiteit der beide genoemde beroepen.

Een dergelijke interpretatie berust echter op een aantal veronderstellingen, die zeker niet onbeperkt geldig zijn. In de eerste plaats wordt verondersteld, dat onderzochte en onderzoeker bij benadering dezelfde visie hebben op de betreffende beroepen, hetgeen niet altijd opgaat. De meeste mensen hebben van het werk van een bakker nog wel enige notie, al bestaan reeds hieromtrent verrassende misvattingen. Maar ten aanzien van de bankwerker zijn de verschillen in opvatting al veel groter. We moeten er op verdacht zijn, dat de beroepsnaam niet ondubbelzinnig is; bankwerker is niet alleen een metaalbewerker, maar ook een beroepsnaam in de machinale brood- en banketfabricage. Het is niet bij voorbaat uitgesloten, dat de onderzochte bij het invullen van de lijst aan deze tweede betekenis heeft gedacht. Ook verwarring met het beroep van bankemployé is mogelijk en zelfs niet zeldzaam. Verder zijn verschillen in inzicht over het niveau van het beroep gemakkelijk mogelijk. In de praktijk wordt het beroep van metaal-bankwerker uitgeoefend op alle niveau's, die in de indeling van het Rijksarbidsbureau ⁽¹¹¹⁾ in de klassen 1 à 2 t/m 5 worden ingedeeld.

Niet alleen de objectieve zijde van het beroep, ook de subjectieve aspecten kunnen aanleiding geven tot verschillen in opvatting. Op welke aspecten van het beroep is de voorkeur of afkeer gebaseerd? In beginsel zou men kunnen trachten, zowel de objectieve als de subjectieve visie van de cliënt op het beroep te leren kennen, door een appèl te doen op de introspectie van de onderzochte. Dit heeft echter in de praktijk slechts zelden

succes: de gemiddelde mens is niet gewend zijn innerlijke reacties diepgaand te analyseren; evenmin kan hij zijn gevoelens goed onder woorden brengen.

Het voorgaande geldt reeds voor de beoordeling van algemeen bekende beroepen door volwassenen. In de meeste beroepenlijsten komen echter ook beroepen voor, die aan vele mensen slechts vaag of niet bekend zijn. Voor jeugdigen, met hun in het algemeen geringere beroepenkennis geldt dit nog sterker, zodat te eerder misverstanden zullen ontstaan. Bovendien is van jeugdigen welbekend, dat zij in de perceptie van een bekend beroep dikwijls andere aspecten centraal stellen, dan de meeste volwassenen doen ⁽¹⁰⁸⁾. Al deze overwegingen vormen aanleiding ook de tekeninterpretatie van de beroepsvoorkeur met grote omzichtigheid te hanteren. In de constructie van de gangbare beroepenlijsten is deze problematiek echter niet voorzien.

Aan het gebruik van beroepenlijsten zijn echter nog meer bezwaren verbonden. Deze betreffen het principe van de vrije keuze. In de meeste, in ons land gebruikte beroepenlijsten pleegt men de onderzochte vrij te laten in het aantal beroepen ten opzichte waarvan hij hetzij een voorkeur, hetzij een afkeer te kennen geeft. De wijze waarop de op deze wijze gegeven vrijheid wordt opgevat, blijkt echter individueel uiterst verschillend te zijn. In principe bestaat de mogelijkheid van deze verschillen testmethodisch gebruik te maken, in de praktijk geschiedt dit echter niet. Ten gevolge daarvan is het niet mogelijk, een onderlinge vergelijking van verschillende onderzochten te maken, m.a.w., de beroepenlijst mist het fundamentele kenmerk, waarop alle metende tests gebaseerd zijn. Dit houdt in, dat een dergelijke methode hoogstens waarde heeft als hulpmiddel bij een interview, of als steun bij de klinisch-intuïtieve vorming van een persoonlijkheidsbeeld. Hoezeer een gebruik voor deze doeleinden misschien ook te rechtvaardigen valt, een test in eigenlijke zin is de beroepenlijst dus niet.

Een iets gunstiger situatie ontstaat, indien een meer gestandaardiseerde methode wordt gebruikt, door toepassing van de L I D-techniek. Toch verdient ook deze geen krachtige aanbeveling. Uit een onderzoek van PERRY ⁽¹⁰⁶⁾ is gebleken, dat het L I D-principe minder efficiënt is, dan dat van de gedwongen keuze, waarbij een vast aantal + en — reacties moet worden gegeven. Speciaal in het kader van een tekeninterpretatie, waarbij men de voorkeursreacties onderling zou willen vergelijken, geeft de L I D-techniek weinig houvast. Wel zijn er weer pogingen gedaan, de mogelijkheid tot onderlinge vergelijking groter te maken, door de beoordeling te doen plaatsvinden op een schaal met meer dan drie schaalpunten, b.v. op de schaal L! L I D D!, een systeem, dat reeds in de twintiger jaren in de *Carnegie Interest Inventory*, maar later ook zowel in Amerika ⁽²³⁾ als in Europa ^(46, 108) is gebruikt. Dit systeem heeft echter weer het bezwaar, dat het in vrij sterke mate een beroep doet op de introspectie, terwijl boven-

dien de individuele geneigdheid, gebruik te maken van de geboden mogelijkheid tot het geven van extreme reacties, zeer verschillend is. Hoewel men hiertegenover opnieuw zou kunnen stellen, dat dit verschil eventueel zelfs een extra informatie kan opleveren, mits het testmethodisch goed wordt uitgebuit, moet toch worden geconstateerd, dat het bij de huidige stand van de testtechniek alleen als een bezwaar kan gelden.

De conclusie van het voorgaande kan nauwelijks anders zijn, dan dat het gebruik van beroepenlijsten van de beschreven — in ons land nog altijd zeer gangbare — vorm, van testtechnisch standpunt niet verantwoord is. Dit houdt echter niet in, dat alle interesstests, waarin op enigerlei wijze gebruik wordt gemaakt van items met beroepennamen, daarmee ook als ondeugdelijk worden bestempeld. Er zijn integendeel enkele goede pogingen ondernomen, beroepenlijsten te construeren, die niet aan de besproken bezwaren lijden. De belangrijkste hiervan is de *STRONG Vocational Interest Blank*, die afzonderlijk zal worden besproken.

Daarnaast zijn uit ons eigen land enkele, niet-gepubliceerde pogingen bekend geworden, die gebaseerd zijn op het streven uit de beroepenlijst een test met innerlijke structuur te maken. De hierboven beschreven beroepenlijsten bezitten een dergelijke structuur niet, zij bestaan uit een verzameling betrekkelijk willekeurig gekozen beroepen, die als onderling onafhankelijk worden behandeld. Men kan nu trachten een structuur aan te brengen, door de beroepen niet als op zichzelf staande eenheden te behandelen, maar hen integendeel op te vatten als representanten van bepaalde beroepsgroepen of beroepenfamilies ⁽¹¹¹⁾. De beroepenlijst wordt dan geconstrueerd door uit te gaan van deze grotere eenheden, die in bepaalde aantallen in de testitems vertegenwoordigd worden. De beroepen dienen als items. Vanzelfsprekend mag men niet veronderstellen, dat iedere onderzochte de beroepen steeds zal beoordelen onder het aspect van de verwantschap met een grotere categorie. Wel echter mag men de tendentie daartoe aanwezig veronderstellen, wanneer betrekkelijk veel items ter beoordeling worden aangeboden. Een voorkeur voor één concreet agrarisch beroep mag b.v. niet zonder meer worden geïnterpreteerd als een affiniteit tot agrarisch werk; maar een frequente sympathie voor beroepen uit de agrarische sector wijst wel met een vrij grote mate van waarschijnlijkheid op een dergelijke affiniteit. De voorkeursreacties worden dan dus niet meer op het concrete beroep betrokken geïnterpreteerd, maar alleen op de structuur, die aan de test ten grondslag ligt.

Dit systeem kan nog verder worden uitgebouwd, door ook het principe van de gedwongen keuze in te voeren. Men presenteert dan telkens een aantal beroepen, gekozen als representanten van bepaalde gegeneraliseerde beroepsinteressen en laat hieruit een vastgesteld aantal als sympathiek of antipathiek aanwijzen. In principe kan een dergelijke methode, bij zorgvuldige uitwerking, een test leveren, die goede kwaliteiten bezit. Zover bekend is echter tot dusver, althans in ons land, met deze uitwerking

nog geen begin gemaakt. Van de methode bestaat dan alleen een geraamte, dat als zodanig niet veel beter is, dan de te voren besproken beroepenlijsten.

Van het in ons land verrichte onderzoek over beroepenlijsten verdient tenslotte nog een — eveneens niet gepubliceerde — studie van VAN EEKLEN⁽⁴⁶⁾ vermelding. Hij heeft een beroepenlijst ontworpen, die bestaat uit 22 groepen van telkens vijf verwante beroepen. De aard van de verwantschap kan zeer uiteenlopend zijn, zowel groeperingen op grond van technologische als van dieptepsychologische criteria komen voor. Gevraagd wordt steeds de voorkeur voor de items uit te drukken op een vijfdelige schaal (+ + + 0 — — —), deze voorkeur wordt opgevat als symbool voor affiniteit tot het beoordeelde beroep. Daarnaast is door hem een zeer uitvoerige motievenlijst opgesteld. Na de invulling van de beroepenlijst moet de onderzochte op deze tweede lijst aangeven, in hoeverre de genoemde motieven voor hem, in het kader van de waardering der beroepen, betekenis hebben. Ook hiervoor wordt een vijfdelige schaal gebruikt, ditmaal echter gepresenteerd als + + + + + 0 —. In principe kan zo de introspectieve exploratie van de motieven worden vergemakkelijkt en daardoor de interpretatie van de beroepsvoorkeur vereenvoudigd. Ervaring met dit instrument is tot dusver alleen aanwezig ten aanzien van enkele groepen personen met middelbare schoolopleiding. Bij hen konden significante groepsverschillen in de voorkeur en de opgegeven motieven worden geconstateerd. Of de methode ook principieel en praktisch bruikbaar is voor individuele diagnostiek is nog niet bekend. Het zou echter wenselijk zijn, dat hieraan een nadere studie werd gewijd.

STRONG VOCATIONAL INTEREST BLANK

Beschrijving: De test heeft afzonderlijke modellen voor mannen en vrouwen^(122, 123). Ieder van hen bestaat uit 400 items, verdeeld over acht rubrieken. Gevraagd wordt naar de voorkeur voor 100 beroepen, 36 schoolvakken, 49 vormen van ontspanning, 48 vormen van bezigheid en 47 menselijke eigenschappen (lichamelijk, psychisch en sociaal). In al deze gevallen geschiedt de beoordeling volgens het LID-principe. In de beide volgende rubrieken moet een gedwongen keuze worden gemaakt. Uit tien vormen van bezigheid, tien bevredigingsmomenten van de arbeid, tien personen die een bepaald aspect van de arbeid vertegenwoordigen en tien functies in organisatorisch verband, moeten er telkens drie als de meest sympathieke en drie als de meest antipathieke worden aangewezen. Vervolgens moet uit 40 paren items, van gemêleerde aard, telkens degene worden aangewezen, die de voorkeur heeft. De laatste rubriek bestaat uit 40 vragen, die een zelfbeoordeling eisen, hier geldt weer het beginsel van de vrije keuze.

De scoring van deze test geschiedt afzonderlijk voor ieder beroep of iedere richting, waarvoor men de belangstelling zou willen bepalen. Ieder antwoord op iedere item krijgt een gewogen score, die verschillend is naar de aard van het beroep waarvoor de belangstelling wordt gepeild. Deze gewogen scores,

die kunnen variëren tussen $+4$ en -4 , worden algebraïsch gesommeerd en de totaalscore wordt vergeleken met die van mensen, die in het betreffende beroep een vooraanstaande positie innemen. Dit houdt dus in, dat de test zoveel malen moet worden gescoord, als er beroepen zijn, waarvoor men de belangstelling zou willen bepalen.

De STRONG-test is de enige belangstellingstest, waarover ook handboeken zijn verschenen (⁴¹, ¹²⁴). Over het geheel heeft deze test een uitstekende pers, omdat de validiteitsonderzoekingen in vele gevallen tot zeer positieve resultaten hebben geleid. Grote indruk heeft b.v. het feit gemaakt, dat deze test in het bekende onderzoek naar de geschiktheid voor opleiding tot klinisch psycholoog een grotere validiteit bleek te bezitten, dan enige andere test (⁷⁵). Toch heeft ook deze methode schaduwzijden. De tijdsduur is aan de lange kant. Volgens STRONG (¹²¹) zouden de 400 items door een volwassen man gemiddeld in een half uur kunnen worden beantwoord, terwijl ongeveer 15-jarige jongens gemiddeld ca. 36 minuten nodig zouden hebben. SUPER (¹²⁶, pag. 411) geeft echter aan, dat begaafde volwassenen ruim een half uur en minder goed begaafde of minder stabiele personen meer dan een uur nodig zouden hebben. Deze laatste tijdschatting ligt meer in de lijn van onze eigen ervaringen. Enkele proeven met een voorlopige Nederlandse vertaling van de STRONG-test leidden tot de bevinding, dat een aanzienlijk deel van de leerlingen der zesde klassen twee of meer uren nodig heeft voor het voltooien der test. Welbekend is trouwens, dat Nederlandse proefpersonen over het algemeen voor het invullen van „inventories”, vragenlijsten e.d. meer tijd behoeven, dan voor Amerikanen uit een vergelijkbare bevolkingscategorie wordt vermeld.

Bij een test, waarvan de totale tijdsduur wordt bepaald door de traagste personen uit een groep, is een gemiddelde tijdsduur van meer dan twee uur vanzelfsprekend een ernstig praktisch bezwaar.

Een tweede bezwaar is, dat de test vrij hoge eisen stelt aan het intellectueel niveau van de onderzochten. Dit zal wel moeten worden verklaard uit het feit, dat de test is ontwikkeld in een universitair milieu, waar men vooral in de belangstellingsstructuur van de studenten was geïnteresseerd. Van sommige voorlopers van de STRONG-test wordt expliciet vermeld, dat zij bestemd waren voor het *college*-niveau. Dit maakt ook begrijpelijk, dat de meeste schalen, die tot nu toe voor deze test zijn geconstrueerd, bestemd zijn voor intellectuele, of althans hoofdarbeidersberoepen. Voor de gemiddelde leerling van onze zesde klassen, is de test rijkelijk moeilijk. Daarbij moeten wij er evenwel, ter vermijding van misverstand, op wijzen, dat STRONG de test ook niet voor een zo jonge leeftijdsgroep heeft ontworpen.

Het voornaamste bezwaar tegen toepassing van de test is echter gelegen in de wijze van scoring. Deze moet immers voor ieder beroep of iedere beroepsrichting afzonderlijk geschieden. Wordt de scoring met handkracht gedaan, dan vergt dit, zelfs wanneer men van schablonen gebruik maakt,

tien à vijftien minuten per schaal. Indien men slechts op één of twee schalen wil beoordelen, zou een dergelijke tijdsduur nog acceptabel kunnen zijn. Maar gewoonlijk moet men in veel meer opzichten beoordelen. De scoring op b.v. een tiental schalen zou echter reeds per onderzochte ongeveer twee uur kosten. Voor toepassing voor massaal onderzoek is dit natuurlijk een onoverkomelijk bezwaar. In de Verenigde Staten zijn de testformulieren daarom ingericht voor scoring met behulp van IBM-apparatuur. De formulieren worden dan opgezonden aan een centrum, dat over een dergelijke installatie beschikt. Zelfs in de Verenigde Staten worden echter de kosten van deze procedure — terecht of ten onrechte — soms als een bezwaar gevoeld (¹²⁶, pag. 413). Voor Nederland, waar door het verschil in levensstandaard deze kosten relatief nog veel hoger zouden zijn, zijn zij prohibitief. Vooral om deze reden komt daarom een bewerking van de STRONG Vocational Interest Blank voor toepassing bij de school- en beroepskeuze in ons land, vooralsnog niet in aanmerking.

K U D E R P R E F E R E N C E R E C O R D

In tegenstelling tot de STRONG-test, die bestaat uit items, welke betrekkelijk willekeurig, d.w.z. niet gebaseerd op een referentiekader, zijn gekozen, bezit de KUDER-test een duidelijke innerlijke structuur. KUDER is uitgegaan van een aantal, op aprioristische overwegingen gekozen interessenrichtingen, waarvoor hij de belangstelling met behulp van zijn test wilde peilen. In enkele gevallen (sport, godsdienst, sociaal prestige) is het niet gelukt een voldoende homogene schaal te construeren, deze richtingen zijn niet in het uiteindelijke model van de test opgenomen. In zijn tegenwoordige vorm is de test gebaseerd op tien belangstellingsrichtingen: buitenwerk, techniek, werken met getallen, (natuur)wetenschap, overreding, beeldende kunst, muziek, verbaal werk, maatschappelijk werk, administratie.

Beschrijving: De items van de test zijn ditmaal geen beroepen, maar activiteiten, die in verband staan met de genoemde belangstellingsrichtingen en daardoor dikwijls ook met het beroepsleven. Door homogeniteitsanalyse is nagegaan, tot welke schaal of schalen iedere item behoort. De items zijn verzameld in groepen van drie, die tot verschillende schalen behoren. Uit ieder van deze triaden moet één item als de meest sympathieke en één als de minst sympathieke worden aangewezen. KUDER past dus het principe van de gedwongen keuze toe. In totaal moeten op deze wijze 168 triaden worden beoordeeld. De scoring kan met de hand geschieden omdat de test slechts éénmaal hoeft te worden beoordeeld. Dank zij een zeer ingenieuze, maar druktechnisch vrij kostbare uitvoering van het antwoordformulier kan dit op snelle wijze geschieden. Behalve de scores op de tien schalen verkrijgt men ook nog een verificatiescore, die vooral dient als controle op mogelijke afwijkingen van de testinstructie.

Behalve de gewone, „vocational” vorm van zijn test ⁽⁸³⁾, heeft KUDER nog twee andere modellen van de *Preference Record* geconstrueerd. De „personal” ⁽⁸²⁾ bestaat eveneens uit 168 triaden, die ditmaal echter tot vijf sectoren behoren: werken in groepsverband, voorkeur voor vertrouwde gelijkblijvende situaties, werken met ideeën, conflicten vermijden, leiding geven. De scoring geschiedt geheel analoog aan die van de „vocational” vorm.

De recente „occupational” vorm ⁽⁸⁰⁾ bestaat uit 100 triaden van activiteiten, die van dezelfde aard zijn als die van de eerstgenoemde vorm. Ditmaal wordt echter bij de scoring op dezelfde wijze te werk gegaan, als bij de STRONG Vocational Interest Blank. Deze geschiedt dus per beroep of beroepsrichting, ook weer doordat aan ieder antwoord op ieder item een bepaalde gewichtsfactor wordt toegekend. Het motief voor deze gewijzigde uitwerkingwijze ligt zeer waarschijnlijk in het feit, dat de methode van STRONG tegenover zijn praktische nadelen het voordeel van een hogere validiteit bezit, terwijl de meeste bezwaren ertegen door gebruik van een machinale scoringswijze kunnen worden opgeheven. Tot dusver is de bruikbaarheid van deze test beperkter dan die van STRONG, omdat nog slechts weinig schalen zijn geconstrueerd. KUDER heeft echter de goede gedachte gehad, een „research-handbook” te publiceren, dat een betrekkelijk eenvoudige handleiding geeft voor gebruikers, die zelf verdere schalen zouden willen ontwikkelen ⁽⁸¹⁾.

Het feit, dat er inderdaad mentaliteitsverschillen tussen de gemiddelde Amerikaanse en de gemiddelde Europese psycholoog bestaan, wordt wel getypeerd door het verschijnsel, dat in Europa de KUDER *Preference Record* met zijn duidelijke structuur bij velen beter in de smaak valt, dan de sterk behavioristische STRONG *Vocational Interest Blank*, terwijl voor de Verenigde Staten juist het omgekeerde geldt. Weliswaar geniet ook daar de KUDER-test vrij veel aanzien maar deze staat toch duidelijk lager aangeschreven dan de STRONG-test. Tegen de *Preference Record* wordt onder andere aangevoerd, dat deze te doorzichtig zou zijn en dus zeer kwetsbaar voor falsificatie — een gevaar, dat overigens bij alle belangstellingstests aanwezig is. Een gewichtig bezwaar is voorts, dat er over deze test nog maar weinig validatiestudies bestaan. Het weinige materiaal, dat er is ⁽⁸⁹⁾, wijst er echter wel op, dat de *Preference Record* een goede validiteit bezit, zodat er geen reden is, de test om deze reden geheel af te wijzen. Wel geldt, dat de validiteit groter wordt, naarmate de ijking van de test specifiekier afgestemd wordt op de situatie, die men wil beoordelen. De constructie van schalen voor afzonderlijke beroepen, zoals dit bij de „occupational” vorm is voorzien, zal vermoedelijk een grotere validiteit in de predictie ten opzichte van deze beroepen mogelijk maken. Bij de oorspronkelijke werkwijze moet men immers door een combinatie van bevindingen, betreffende de hoofdrichtingen van de belangstelling, komen tot een beoordelen van het interesse voor specifieke beroepen. Een derge-

lijke beoordeling via een tussenstadium maakt gebruik van een klinische interpretatie. Bij de huidige stand van de wetenschap is deze dikwijls minder valide dan een rechtstreekse beoordeling, gebaseerd op statistische basis. (¹⁵, ⁹⁸, ⁹⁹). Maar evenals bij de STRONG-test geldt ook hier, dat een toepassing op grote schaal in ons land voorlopig op materiële bezwaren moet afstuiten.

In vergelijking met de STRONG *Vocational Interest Blank* heeft de KUDER *Vocational Preference Record* een aantal voordelen. De test is veel minder tijdrovend, gemiddeld zijn alle deelnemers aan een klassikaal onderzoek binnen het bestek van ruim een uur gereed. Verder stelt de test minder hoge eisen aan het begrip van de onderzochte. Uit een Amerikaans onderzoek (¹¹⁴, geciteerd bij ¹²⁶, pag. 446) bleek, dat de vocabulaire van de KUDER-test binnen het bereik van de gemiddelde Amerikaanse achtste klasser lag, terwijl de STRONG-test op het niveau van de gemiddelde tiende klasser kwam. Ook de eerstgenoemde waarde zou de indruk kunnen wekken, dat de test voor gebruik bij selectie aan het einde van ons lager onderwijs nog rijkelijk moeilijk zou zijn. De ervaring leert echter, dat — misschien door de meer schoolse instelling van ons lager onderwijs — de eisen, die een item aan het theoretisch begripsvermogen stelt, bij ons vroeger worden vervuld dan in de Verenigde Staten. Sommige intelligentietests, die in de Verenigde Staten op het niveau van de achtste klasse liggen, kunnen bij ons reeds in de zesde klasse worden gebruikt. Enkele proeven met een vertaling van de KUDER-test hebben geleerd, dat ook deze in een gemiddelde zesde klasse wel bruikbaar zal zijn, alleen voor de zwakke leerlingen zou de test aan de moeilijke kant kunnen zijn. Natuurlijk bevat de test wel items, die voor de gemiddelde twaalfjarige weinigzeggend zijn, maar dit wordt in de ijking automatisch verdisconteerd.

Een belangrijk voordeel van de *Preference Record* is nog, dat deze bezigheden als items gebruikt, in plaats van beroepen. Activiteiten geven in veel geringer mate dan beroepsnamen, liefhebberijen, e.d., aanleiding tot misverstanden. Verder voldoet de test ten aanzien van de betrouwbaarheid aan hoge eisen, de gevonden testbetrouwbaarheids-coëfficiënten (⁷⁹, ¹²⁶ pag. 452) liggen tussen .84 en .98.

Een laatste aangelegenheid betreffende de KUDER *Preference Record* raakt het principe van de gedwongen keuze. Tegen dit principe zijn bezwaren aan te voeren (¹³⁹), die vooral betrekking hebben op het geval, dat slechts een klein aantal keuzen worden gedaan. In dat geval wordt de invloed van toevallige variaties als regel niet voldoende genivelleerd. De sterkte van de voorkeur of afkeer kan van item tot item sterk verschillen, terwijl toch het testprincipe iedere keer tot het geven van dezelfde reactie dwingt. Bij een gering aantal keuzen kunnen deze reacties voor de scoring daarom niet als gelijkwaardig worden behandeld, zij mogen derhalve ook niet zonder meer worden gesommeerd. Is het aantal keuzen echter groot, dan gaat de spreiding der sterkte-verschillen steeds meer een normale

verdeling benaderen, de extreme waarden nivelleren elkaar. We mogen dan de voorkeursreacties wel als gelijkwaardig behandelen en de bezwaren tegen het sommeren van de reacties zijn niet meer geldig. De ervaring heeft geleerd, dat bij 168 en zelfs reeds bij 100 keuzen de nivellerende werking voldoende groot is, om het principe van de gedwongen keuze te mogen toepassen.

De voorgaande overwegingen wijzen er op, dat de *KUDER Preference Record* ernstig in aanmerking zou komen, ten behoeve van het interessen-onderzoek in het kader van de voorlichting bij de overgang van het lager naar het voortgezet onderwijs in Nederland, te worden bewerkt. Het voornaamste praktische bezwaar tegen deze test betreft nog de vrij kostbare uitvoering van het antwoordformulier, waarin een nakijksysteem is verwerkt. Laat men dit laatste weg, dan wordt de scoring een probleem, omdat het antwoordformulier zeer compact moet worden samengesteld. De scoring met behulp van een schablonen vergt dan zeer veel concentratie en geeft dus een aanzienlijke kans op fouten. Door een enigszins andere inrichting van de test zou dit bezwaar echter wel op te heffen zijn, zodat het niet doorslaggevend is. Ook in omliggende landen bestaat dezelfde appreciatie van deze test, getuige de vertalingen, die reeds in verschillende landen zijn gemaakt en het feit, dat de test als model heeft gediend voor de constructie van andere belangstellingstests. Tot deze laatste behoort de Beroepen-Interesse-Test van M. IRLE, die het voorwerp en instrument van de hier gerapporteerde studie is geworden.

2.3. De studie van IRLE

De B I T zal in het volgende hoofdstuk uitvoerig worden beschreven. Ter inleiding zullen echter eerst de werkwijze bij de constructie van de test en het daarop volgende toetsingsonderzoek worden gerapporteerd.

IRLE (⁶⁶, ⁶⁸) is uitgegaan van de volgende principes. Hij wilde een gestructureerde test maken, die voldoet aan de eisen, die bij de huidige stand van de techniek aan homogeniteit, betrouwbaarheid, validiteit worden gesteld. In deze test moesten zoveel mogelijk de belangstellingsrichtingen vertegenwoordigd zijn, die in het kader van de beroepskeuze een rol kunnen spelen. Verder moest de test opgebouwd zijn naar een ook praktisch gemakkelijk te hanteren systeem.

De eerste eis heeft hem vrijwel onvermijdelijk in de richting van een test van het KUDER-type gevoerd, het enige gestructureerde type, dat momenteel aan de genoemde voorwaarden voldoet. Ten aanzien van de keuze van de belangstellingsrichtingen vond hij een aanknopingspunt in een studie van SUPER (¹²⁶, pag. 380 e.v.). Deze heeft een vergelijking gemaakt van de factoren, die in verschillende interessentests een rol spelen. Vergeleken werden de KUDER-test, de uitkomsten van een factoranalyse van de STRONG-test en voorts een tweetal tests, die thans voor het praktisch psychodiagnostisch onderzoek weinig betekenis meer hebben: de

tests van ALLPORT-VERNON en LURIE. Tussen deze tests bleek ten aanzien van de factoren een grote mate van verwantschap te bestaan. Dit is vooral belangrijk, omdat tendele deze factoren niet a priori aan de test ten grondslag zijn gelegd, maar eerst achteraf, factoranalytisch zijn geïsoleerd. Er zijn dus duidelijke aanwijzingen, dat de beroepeninteressen inderdaad volgens een beperkt aantal principes kunnen worden gebundeld. SUPER concludeert tot het bestaan van acht belangstellingsfactoren: scientific, social-welfare, literary, material, system, contact, artistic, musical. Deze indeling heeft IRLE in hoofdzaak overgenomen. De muzikale belangstelling werd weggelaten, hoewel deze als interessefactor zeker reden van bestaan heeft. Dit geschiedde op grond van de praktische overweging dat onderzoek, met behulp van een test, in deze richting als regel overbodig is. De factor „scientific”, die in de Verenigde Staten in hoofdzaak de natuurwetenschappelijk-gerichte belangstelling aanduidt, werd expliciet als natuurwetenschappelijk-technische interesse gedefinieerd, terwijl voor de geesteswetenschap een plaats bij de literaire belangstelling werd gevonden. Deze verschuiving schijnt, in het licht van de Europese opvattingen over geesteswetenschap, wel verantwoord. De factor „system” werd opgevat als administratief interesse, een interpretatie die, gezien de items waarop deze factor in Amerika is gebaseerd, aannemelijk is. Tenslotte werden twee factoren toegevoegd. De eerste in aansluiting aan de belangstelling voor buitenwerk, die in de laatste revisie van de KUDER-test is opgenomen. Deze richting werd hier als belangstelling voor agrarische arbeid geformuleerd. De tweede factor had betrekking op belangstelling voor arbeid in de voedingssector. In totaal komt IRLE daardoor tot negen belangstellingschalen:

1. *Technische handarbeid* („Technisches Handwerk”). Het accent valt in het bijzonder op de metaalbewerking en minder op dat van hout of andere materialen. Naast de eigenlijke bewerking neemt ook de montage een belangrijke plaats in deze schaal in.
2. *Ambachtelijke vormgeving* („Gestaltendes Handwerk”). De schaal onderzoekt vooral de affiniteit tot het kunstambacht, waartoe in dit verband zowel het bewerken als het ontwerpen moeten worden gerekend. Door de beperking tot de toegepaste kunst bestrijkt deze schaal een iets beperkter gebied, dan de verwante schaal voor „artistic interests” uit Amerikaanse tests.
3. *Techniek en natuurwetenschap* („Technische und naturwissenschaftliche Berufe”). In deze schaal komen ook items voor, die betrekking hebben op biologie, meteorologie en andere randgebieden van de natuurwetenschap. Maar de nadruk valt toch sterk op de anorganische grondwetenschappen, natuur- en scheikunde en op de ingenieurswetenschap; de hierop betrekking hebbende items hebben de hoogste homogeniteitscoëfficiënten.

4. *Voedselbereiding* („Ernährungshandwerk”). In deze schaal komen items van zeer verschillende aard voor. Sommige hebben een wetenschappelijke inslag, andere hebben betrekking op de voedselbereiding in engere zin, weer andere verwijzen naar productie-activiteiten met een leidinggevende of op andere wijze verantwoordelijkheid-dragende aard. Deze laatste blijken de hoogste mate van homogeniteit te bereiken, de schaal verwijst dus vooral naar de hogere functies en de zelfstandige beroepen in de voedselbereiding. Opgemerkt moet worden, dat er in deze schaal items voorkomen, die door de meeste onderzochten niet erg au sérieux worden genomen („ossen aan het spit braden”). Dergelijke items hebben een lage homogeniteit. Hun aanwezigheid in de test moet blijikbaar vooral worden verklaard uit het feit, dat de constructie van deze schaal veel meer moeilijkheden heeft gegeven, dan die der andere schalen. Het is dermate moeilijk geweest, een voldoende aantal representatieve items te vinden, dat twee items tweemaal zijn gebruikt. We zullen later zien, dat ook bij ons eigen onderzoek de homogeniteit der items van deze schaal een zwak punt in de test is gebleken; nochtans bezit zij een behoorlijke validiteit.
5. *Agrarische arbeid* („Land- und forstwirtschaftliche Berufe”). Behalve activiteiten die verband houden met land- tuin- of bosbouw, bevat de schaal ook enkele items, die verband houden met visserij. Gebleken is echter, dat deze laatste veel minder homogeen met de schaal zijn dan de items uit de eigenlijke agrarische sector. De Duitse naam, die ook de bosbouw expliciet vermeldt, geeft wel de beste aanduiding voor dit interessegebied.
6. *Handel* („Kaufmännische Berufe”). De analyse van de homogeniteit der items uit deze sector leidt tot de conclusie, dat er geen speciale subgroepen zijn, die een centrale plaats innemen. De homogeniteitscoëfficiënten zijn echter over de gehele lijn lager, dan bij de meeste andere schalen, hetgeen er o.i. op wijst, dat het gebied als geheel geen sterke eenheid vormt. Bij inspectie der items blijkt nog, dat deze aan zeer verschillende sectoren van de handel zijn ontleend: verkoop op pad, detailhandel, verzekeringsbedrijf, administratief-commercieel werk.
7. *Administratief werk* („Verwaltende Berufe”). De items hebben zowel op de bedrijfsadministratie en -organisatie als op de overheidsadministratie betrekking. De homogeniteitsanalyse wijst er echter op, dat de eerstgenoemde categorie de centrale plaats toekomt.
8. *Literaire en geesteswetenschappelijke arbeid* („Literarische und geisteswissenschaftliche Berufe”). De grootste homogeniteit hebben de items, die betrekking hebben op literaire en journalistieke arbeid. De geesteswetenschap, hier voornamelijk door de historie vertegenwoordigd, blijkt iets minder goed met de totaalscore van de schaal te correleren,

maar bereikt toch nog waarden, die zeker niet ongunstig afsteken bij de meeste homogeniteitscoëfficiënten van de schalen 4 en 6.

9. *Sociaal werk en opvoeding* („Sozialpflege und Erziehung”). De items uit deze schaal hebben vrijwel zonder uitzondering een hoge correlatie met de totaalwaarde van de schaal. Blijkbaar zijn dus de beide componenten van de schaal, wat hun appèl aan de belangstelling betreft, zeer nauw aan elkaar verwant, zodat zij terecht als een eenheid worden behandeld.

Als model voor een praktisch gemakkelijk te hanteren systeem, waarmee deze belangstellingsrichtingen kunnen worden getoetst, heeft IRLE niet de eigenlijke KUDER-test gekozen, maar een in wezen verwante test, de *Motivation Indicator* van BALDWIN (14). Deze Amerikaanse test heeft zeer weinig bekendheid verkregen; afgezien van een tijdschriftbespreking en een vermelding in een handboek (7, 26) is er in de literatuur geen aandacht aan besteed. Kenmerkend voor deze test is het gebruik van een ruimtelijk schema, waarbij de items zowel horizontaal als verticaal in rijen worden geordend. Daardoor grenzen telkens vier items ruimtelijk aan elkaar, zodat zij vergeleken kunnen worden. De concrete uitvoeringsvorm van dit principe zal in het volgende hoofdstuk worden toegelicht; verwezen wordt ook naar bijlage V. De vergelijking der items kan uiteraard op verschillende wijzen geschieden. BALDWIN laat in zijn test uit ieder viertal (tetrad) twee items als de meest sympathieke kiezen. IRLE heeft aanvankelijk één positieve en één negatieve reactie laten geven, maar laat in de definitieve uitvoering van zijn test alleen één positieve sympathie per viertal aangeven. Doordat iedere item, die niet aan de rand van het veld is gelegen, in vier tetraden is opgekomen, komt deze in de loop van de test viermaal ter keuze, zodat de score per item tussen 0 en 4 kan variëren. Deze scores worden voor alle items, die tot eenzelfde belangstellingsrichting behoren, gesommeerd. Per blad van de test zijn 81 items opgenomen. Om ook de items, die aan de rand zijn gelegen een even grote kans als de overigen te geven, gekozen te worden, is een extra laatste rij toegevoegd, die identiek is aan de eerste rij, en een extra laatste kolom, die identiek is met de eerste kolom. Het totale veld bestaat daardoor uit 100 items. Op de vier hoekpunten komt hetzelfde item voor, de overige items aan de buitenrand zijn ieder tweemaal vertegenwoordigd, terwijl de items binnen de randen ieder slechts éénmaal aanwezig zijn.

De B.I.T. bestaat uit twee bladzijden, die in totaal 162 items hebben, d.w.z. 18 per schaal. Er wordt dus ook 162 maal gekozen en de maximale score per schaal bedraagt 72 punten. Bij dit laatste moet echter worden bedacht, dat de scores niet geheel onderling onafhankelijk zijn, omdat de som van de scores op de negen schalen steeds 162 moet zijn.

Aan de hand van het testformulier kan de bedoeling van het systeem gemakkelijk worden uitgelegd. Uit onze eigen onderzoekingen is gebleken,

dat leerlingen uit de zesde klasse vrijwel steeds zonder veel moeite de uitleg begrijpen en de testinstructie goed uitvoeren. Het grote voordeel van het schema van BALDWIN is, dat het zich gemakkelijk voor een scoring met behulp van schablonen leent; deze kan snel, zonder grote kans op fouten en weinig kostbaar worden uitgevoerd. Het voornaamste bezwaar tegen dit systeem is van test-constructieve aard: door de ruimtelijke ordening zijn de items niet onderling onafhankelijk. Blijkt bij de itemanalyse een item ondeugdelijk, zodat deze vervangen moet worden, dan heeft dit invloed op de waarde van de aangrenzende items. Zou men in het kader van een revisie van de test een enigszins aanzienlijk aantal items willen vervangen, dan heeft dit tot consequentie, dat de gehele statistische bewerking van de test opnieuw moet worden uitgevoerd. Daarom is het, indien een enkel item ongeschikt blijkt, in de praktijk het eenvoudigst, dit item „stil te leggen”, d.w.z. bij de berekening van de totaalscores niet mee te tellen. Zolang deze procedure tot enkele gevallen beperkt blijft, heeft dit op de gevoeligheid van de betreffende schaal geen noemenswaardige invloed.

IRLE heeft het grondschema van BALDWIN overgenomen, maar daarin uiteraard zijn eigen items geplaatst. De ordening van de items is zodanig, dat nooit twee items uit dezelfde schaal met elkaar behoeven te worden vergeleken. Het grondschema is hierna weergegeven, waarbij de nummers corresponderen met de te voren besproken belangstellingsrichtingen.

1	4	7	3	6	8	2	5	9	1
2	6	9	2	4	9	3	4	8	2
3	5	8	1	5	7	1	6	7	3
1	4	7	3	6	8	2	5	9	1
2	6	9	2	4	9	3	4	8	2
3	5	8	1	5	7	1	6	7	3
1	4	7	3	6	8	2	5	9	1
2	6	9	2	4	9	3	4	8	2
3	5	8	1	5	7	1	6	7	3
1	4	7	3	6	8	2	5	9	1

Op de merites van dit schema wordt door IRLE niet ingegaan. Wij willen in dit verband volstaan, met de aandacht te vestigen op een enkel aspect. Het valt op, dat het grondplan in wezen uit slechts drie rijen bestaat, die driemaal worden herhaald. Dit wijst er reeds op, dat niet alle mogelijke combinaties van vier uit negen schalen zullen worden getoetst. In totaal zijn immers 126 van dergelijke combinaties mogelijk, terwijl er in het hier gegeven schema slechts 25 verschillende voorkomen. Daar IRLE bovendien voor de beide bladen van zijn test hetzelfde schema heeft gebruikt, brengt ook dit geen vermeerdering van het aantal getoetste combinaties mee. Ten aanzien van de combinaties van drie, zoals deze in de tetraden kunnen worden onderscheiden, ligt de situatie reeds wat gunstiger:

van de 84 mogelijke combinaties worden er 54 getoetst. Tenslotte leert een analyse van het schema, dat alle 36 mogelijke combinaties van twee schalen met elkaar worden vergeleken, zij het ook niet alle even vaak. Er zijn negen combinaties van twee schalen, die per bladzijde 18 maal worden vergeleken, twee combinaties, die 15 maal, 23 combinaties, die 12 maal en 2 combinaties, die 9 maal aan elkaar worden getoetst. In theorie zou dit een bezwaar kunnen zijn, omdat een schaal, die bij een bepaalde categorie personen een sterk appèl heeft en dus vaak wordt gekozen, een zeker „wegdruk-effect” uitoeft, d.w.z. maakt dat de schalen, waarmee hij wordt vergeleken, minder vaak worden gekozen. Het zou nu denkbaar zijn, dat een schaal, die zeer frequent met een dergelijke populaire schaal wordt vergeleken, een merkbaar lagere gemiddelde score zou krijgen. Bij het statistisch onderzoek van de test is echter niet gebleken, dat dit verschijnsel reële betekenis heeft, zodat het genoemde bezwaar praktisch niet geldig is.

Bij de beoordeling van het bovenstaande moet men bedenken, dat door de ruimtelijke ordening de vrijheid tot groeperen van de items wordt ingeperkt; er mogen immers geen twee items, die tot dezelfde schaal behoren, aan elkaar grenzen. Het topologische probleem, of door een andere groepering van de items een aanmerkelijk groter aantal combinaties van drie en vier kan worden getoetst, is door ons niet nader onderzocht. Voor de hand ligt echter wel, dat gemakkelijk een vermeerdering van het aantal getoetste combinaties had kunnen worden bereikt, door aan de achterzijde van het testformulier de negen gebieden op een andere wijze te nummeren. IRLE heeft dit niet gedaan, om dezelfde nakijkschablonen voor voor- en achterzijde van de test te kunnen gebruiken, hetgeen een niet te onderschatten praktisch voordeel oplevert.

Bij de constructie van de test is IRLE uitgegaan van het dubbele aantal items, d.w.z. 36 per schaal. Deze waren op vier bladen gegroepeerd. Met behulp van een nomografische methode ⁽⁵⁹⁾ werd van alle items de tetrachorische correlatie met de schaal, waarin zij op grond van apriorische overwegingen werden ingedeeld, berekend. Uit de publicatie van IRLE blijkt niet duidelijk, hoe de grenslijnen voor deze bepaling zijn getrokken, het waarschijnlijkst is, dat de berekening op de medianen of de gemiddelden is geschied. Een dergelijke correlatieberekening dient ter bepaling van de homogeniteit der items met de bijbehorende schaal. Uit iedere schaal werden de 18 items met de laagste correlaties geëlimineerd. In één geval, bij de voedselbereiding, bleken ook de overige correlaties meest zo laag, dat het nodig was, nieuwe items te construeren. Daarna werd met de overgebleven 162 items een nieuw model van de test samengesteld, dat ditmaal uit twee bladen bestond. Aangezien hierbij een hergroepering van de items plaatsvond, was het noodzakelijk, de correlatieberekeningen te herhalen. De uitkomsten van deze berekeningen zijn op een gemakkelijk toegankelijke plaats gepubliceerd ⁽⁶⁷⁾, pag. 11 e.v.), we geven hier alleen,

in tabel 2.1, een samenvatting van de extreme en de mediale waarden per schaal. De schalen zijn afgekort aangeduid met twee letters, die aan hun naam zijn ontleend.

TABEL 2.1. Variatiegrenzen van de homogeniteit der items, volgens IRLE.

Schaal	Variatiebreedte	Mediaan
TH	.40 — .95	.85
AV	.49 — .93	.75
TN	.49 — .95	.76
VB	.12 — .80	.57
AA	.54 — .90	.72
HA	.43 — .90	.55
AD	.51 — .83	.68
LG	.42 — .80	.67
SO	.64 — .90	.78

Overgenomen uit ⁶⁸, pag. 33.

Volgens IRLE zouden deze nomografisch bepaalde correlaties over de gehele lijn wat lager liggen dan de algebraïsch berekende. In de publicatie over de methode (⁶⁹) is hiervan echter geen aanduiding te vinden. De zeer hoge correlaties moeten met enige scepticisme worden bekeken, aangezien de dichotomie hierbij waarschijnlijk zeer extreem is geweest. In ieder geval blijkt echter uit de tabel wel, dat de grote meerderheid van de items inderdaad een goede homogeniteit met de schaal waartoe zij worden gerekend, bezit.

De test is door IRLE vervolgens geijkt op een overwegend manlijke proefgroep van ruim 1400 personen (1150 mannen, 260 vrouwen). De leeftijdsverdeling wordt niet nauwkeurig meegedeeld; men kan echter schatten, dat de mediale leeftijd 16 à 17 jaar bedroeg, terwijl de minimum-waarde ca. 14 jaar was en het maximum op volwassen leeftijd lag. Het aantal handarbeiders, of jeugdigen die later een handarbeidersberoep gaan uitoefenen, overweegt in het materiaal sterk. Er is geen poging gedaan, te onderzoeken of de ijkingsgroep bij benadering een representatieve steekproef, b.v. uit de beroepsbevolking of uit een bepaalde leeftijdssector van de bevolking vormt. De ijkingsgegevens zijn in eerste instantie omgerekend op percentielscores, die vervolgens grafisch genormaliseerd zijn (⁶⁷, pag. 10).

Ter controle van de statistische kwaliteiten van de test heeft IRLE de betrouwbaarheidscoëfficiënten van de negen schalen berekend, met behulp van één der formules van KUDER-RICHARDSON. De gevonden correlaties variëren tussen .811 (voedselbereiding) en .949 (agrarische arbeid), zij zijn dus voldoende hoog voor individuele toepassing van de test.

Verder is een onderzoek ingesteld naar de intercorrelaties tussen de schalen. Deze worden in tabel 2.2 weergegeven.

TABEL 2.2. Intercorrelaties der BIT-schalen volgens IRLE.

	AV	TN	VB	AA	HA	AD	LG	SO
TH	— .38	.43	— .25	— .07	— .10	— .15	— .30	— .41
AV		— .49	.17	.10	— .07	— .35	.31	.02
TN			— .36	— .13	— .31	.11	— .14	— .42
VB				.05	— .08	— .24	— .22	.05
AA					— .61	— .53	.03	— .07
HA						.24	.07	— .06
AD							— .19	— .06
LG								.05

Bij inspectie van deze tabel blijkt, dat de schalen TH en TN en ook HA en AD zich duidelijk van de overige onderscheiden. „Spiegeling” van de betreffende coëfficiënten maakt de grote meerderheid der correlaties positief.

IRLE wijst er op, dat de hier gevonden correlaties niet zonder meer vergelijkbaar zijn met de intercorrelaties van de subtests van b.v. een intelligentietest. Door de vastgestelde totaalscore van 162 zijn de scores op de schalen onderling afhankelijk. Daardoor mogen ook de correlatiecoëfficiënten niet als onderling onafhankelijke grootheden worden beschouwd, hetgeen consequenties voor de interpretatie kan hebben. Veel betekenis heeft deze onderlinge afhankelijkheid vermoedelijk niet. Men kan aantonen, dat in het geval van een aselechte beantwoording van de test de onderlinge correlatie tussen de schalen gelijk is aan $-m/216$, waarin m het aantal malen voorstelt, dat items uit de twee te correleren schalen op één bladzijde van de test met elkaar worden vergeleken. Wij hebben hiervoor gezien, dat het aantal combinaties 2 aan 2 van twee schalen per bladzijde varieert tussen 9 en 18. Door substitutie van deze waarden in bovenstaande formule volgt, dat de toevalscorrelatie van de schalen ligt tussen $-.042$ en $-.083$, hetgeen dus zeer gering is. De mate van de onderlinge afhankelijkheid bij niet-aselechte beantwoording is moeilijker te berekenen, maar het ligt niet voor de hand, dat de grootte-orde van het effect aanmerkelijk anders zal worden. Wij zijn daarom geneigd, het standpunt te verdedigen, dat men in eerste instantie bij de interpretatie van de correlatiecoëfficiënten mag handelen, alsof deze niet beïnvloed worden door het feit, dat de totaalscore op de test vastligt.

Door deze vooronderzoekingen kon IRLE aantonen, dat zijn methode aan de internationaal gestelde eisen van testconstructieve aard voldoet. Hij heeft vervolgens een onderzoek ingesteld naar de prestaties van verschillende categorieën beroepsbeoefenaars. In verband met het feit, dat het praktisch bezwaarlijk was, voldoende personen, die reeds in het beroep staan, voor onderzoek ter beschikking te krijgen, heeft hij zijn studie gericht op leerlingen, die b.v. in het kader van hun bezoek aan een beroepsschool gemakkelijker konden worden benaderd. Teneinde enige zekerheid

te verkrijgen, dat hun interesses inderdaad specifiek voor hun beroep waren, zijn alleen personen in het onderzoek betrokken, die bij navraag dit beroep ook aanwezen als hun „lievelingsberoep”. Met uitzondering van drie categorieën, bedroeg de gemiddelde leeftijd in alle onderzochte beroepsgroepen ongeveer 16 jaar. De drie uitzonderingsgevallen betroffen: „Refa-Anwärter” (ongeveer onze aspirant-tijdwaarnemers en -calculators), „Regierungsinspektoren-Anwärter des Technischen Dienstes” (aspirant technisch opzichter) en „Regierungsinspektoren-Anwärter des Allgemeinen Dienstes” (leerling administratief ambtenaar), die gemiddeld respectievelijk 32, 28 en 21 jaar waren. De grootte van de onderzochte groepen varieerde tussen 11 en 106 personen, met een mediale waarde van 35. Van iedere groep werd van alle schalen de gemiddelde score bepaald. Alle groepen werden ook 2 aan 2 vergeleken, om na te gaan op welke schalen een significant verschil tussen de gemiddelden bestond.

In het onderzoek werd nog een laatste groep betrokken, bestaande uit huishoudelijk personeel van zeer eenvoudig niveau. De betreffende meisjes onderscheidden zich van alle overige proefpersonen, doordat zij géén positieve toewending tot het uitgeoefende beroep beweerden te hebben. Hun preferentie ging steeds naar een ander beroep (kleuteronderwijzeres, kraamverzorgster, kapster, naaister of verkoopster) uit.

Bij het onderzoek bleek, dat er vrijwel steeds t.a.v. de gemiddelde preferenties van de beroepsbeoefenaren tussen de beroepen significante verschillen bestonden. In de meeste gevallen hadden deze zelfs op meerdere schalen betrekking. Slechts in één geval: tussen leerling-technisch tekenaar en leerling-timmerman bleek geen enkel verschil significant. Verder waren in een aantal gevallen, van erkend verwante beroepen, de significanties van betrekkelijk bescheiden aard ($.05 > P > .01$). Belangrijker nog dan deze constatering, was de bevinding, dat de gevonden verschillen ook inzichtelijk meestal direct aanvaardbaar waren. Dit blijkt ook uit tabel 2.3, waarin voor iedere belangstellingsrichting een opsomming van de 26 beroepen, die in het onderzoek waren betrokken, is gegeven, in volgorde van de hoogte der gemiddelde score op de betreffende schaal. De aanduiding „leerling” is korthedshalve steeds weggelaten.

Van belang zijn uiteraard vooral de beroepen, die in deze rangorden de hoogste of de laagste plaatsen innemen. De rangorde in het tussenliggend gebied is van veel geringer betekenis, omdat hier de onderlinge verschillen veel kleiner zijn en dus ook gemakkelijk door toevallige factoren te beïnvloeden. Bij analyse van de extreme scores zien we het volgende:

In schaal TH worden de vijf hoogste plaatsen bezet door de vijf metaalbewerkersberoepen, hetgeen geheel in de lijn der verwachtingen op grond van de homogeniteitsanalyse ligt. Bij de bezetters van de laagste plaatsen vinden we de beide vrouwelijke beroepen en voorts een aantal specifiek niet-technische beroepen. Alleen de lage plaats van de drukkers is on-

TABEL 2.3 Rangorde van de beoefenaars van 26 beroepen naar hun gemiddelde schaalscores (volgens IRLB)

<i>Techn. handarbeid</i>	<i>Amb. vorm- geving</i>	<i>Techn. & Nat. Wet.</i>	<i>Voedsel- bereiding</i>	<i>Agrarische arbeid</i>	<i>Handel</i>	<i>Admin. werk</i>	<i>Lit. & geestes- wet.</i>	<i>Soc. werk & opv.</i>
fijnbankwerker	naaister	techn. opzichter	kok	bosopzichter	commerc. empl.	gemeente ambt.	drukker	huishoudhulp
automonteur	boekbinder	electriciën	banketbakker	tuinman	gemeente ambt.	commerc. empl.	gemeente ambt.	naaister
bankwerker	huishoudhulp	fijnbankwerker	kellner	molenaar	adm. ambtenaar	adm. ambtenaar	adm. ambtenaar	gemeente ambt.
loodgieter	meubelmaker	tijdwaarnemer	slager	slager	meetselaar	tijdwaarnemer	commerc. empl.	commerc. empl.
electriciën	banketbakker	automonteur	bakker	bakker	automonteur	drukker	kellner	tijdwaarnemer
molenaar	kok	loodgieter	naaister	boekbinder	bakker	techn. opzichter	naaister	slager
meubelmaker	tuinman	timmerman	meetselaar	naaister	huishoudhulp	timmerman	huishoudhulp	bakker
meetselaar	kellner	bosopzichter	huishoudhulp	meubelmaker	kellner	bankwerker	molenaar	kok
tijdwaarnemer	drukker	bankwerker	tuinman	banketbakker	slager	tekenaar	kok	boekbinder
tekenaar	tekenaar	drukker	molenaar	kok	meubelmaker	kellner	banketbakker	adm. ambtenaar
bakker	timmerman	boekbinder	automonteur	meetselaar	bankwerker	meetselaar	bakker	tuinman
slager	techn. opzichter	adm. ambtenaar	boekbinder	loodgieter	tijdwaarnemer	huishoudhulp	tekenaar	banketbakker
techn. opzichter	meetselaar	bakker	fijnbankwerker	huishoudhulp	timmerman	automonteur	meetselaar	meetselaar
boekbinder	loodgieter	molenaar	meubelmaker	tekenaar	electriciën	electriciën	bankwerker	bankwerker
timmerman	bakker	gemeente ambt.	timmerman	timmerman	boekbinder	loodgieter	slager	electriciën
tuinman	commerc. empl.	commerc. empl.	electriciën	techn. opzichter	loodgieter	molenaar	tuinman	loodgieter
kellner	slager	tuinman	loodgieter	electriciën	banketbakker	slager	loodgieter	timmerman
bosopzichter	adm. ambtenaar	tekenaar	bankwerker	kellner	molenaar	banketbakker	bosopzichter	drukker
banketbakker	gemeente ambt.	meetselaar	techn. opzichter	fijnbankwerker	naaister	naaister	electriciën	kellner
adm. ambtenaar	fijnbankwerker	kellner	drukker	automonteur	drukker	kok	timmerman	bosopzichter
commerc. empl.	bosopzichter	meubelmaker	tekenaar	bankwerker	tekenaar	boekbinder	boekbinder	techn. opzichter
drukker	automonteur	banketbakker	gemeente ambt.	drukker	kok	bakker	fijnbankwerker	meubelmaker
kok	electriciën	slager	bosopzichter	adm. ambtenaar	fijnbankwerker	fijnbankwerker	meubelmaker	automonteur
huishoudhulp	molenaar	kok	adm. ambtenaar	tijdwaarnemer	techn. opzichter	bosopzichter	techn. opzichter	molenaar
gemeente ambt.	bankwerker	huishoudhulp	tijdwaarnemer	gemeente ambt.	tuinman	tuinman	tijdwaarnemer	tekenaar
naaister	tijdwaarnemer	naaister	commerc. empl.	commerc. empl.	bosopzichter	meubelmaker	automonteur	fijnbankwerker

verwacht. Zij moeten wel met machines omgaan, terwijl op grond van een psychologische analyse moet worden aangenomen, dat zij arbeid met een technische inslag verrichten (5, 120). Gezien het feit, dat dezelfde drukkers de eerste plaats in de rangorde bij schaal LG bezetten, hetgeen eveneens in strijd met de verwachting is, zou men zich kunnen afvragen, of IRLE onder de naam „Buchdrucker” wellicht ook letterzetters heeft gerekend. Op deze kwestie zullen wij in hoofdstuk X, naar aanleiding van ons eigen onderzoek, terugkomen.

Ten aanzien van schaal AV valt op, dat de beide vrouwelijke beroepen nu in de top voorkomen. Voor de naaisters zou dit nog in de lijn van de verwachtingen kunnen liggen, voor de huishoudhulpen echter niet. Hier speelt echter ook het sexeverschil een rol, bij ons eigen onderzoek is gebleken, dat deze schaal door meisjes gemiddeld veel vaker wordt gekozen, dan door jongens. Bij de manlijke beroepen die in de tabel boven aan staan, ligt de verwantschap met het artistiek-creatieve voor de hand. De beroepen onder aan de ranglijst staan er omgekeerd inderdaad voor bekend, dat zij aan artistieke neigingen bijzonder weinig bevredigingsmogelijkheid kunnen bieden. Opvallend is hier dat de beide „agrarische” beroepen, tuinman en bosopzichter op de ranglijst ver uiteenstaan. De richting van het verschil dekt zich echter met de overweging, dat in het beroep van tuinman een bescheiden artistieke uitingsmogelijkheid aanwezig is.

Bij schaal TN komen de technische beroepen weer naar boven. Ditmaal bevinden zich ook de beide beroepen met een intellectuele inslag in de hoogste laag. Opvallend laag zijn de technische tekenaars geplaatst, die echter bij het gehele onderzoek van IRLE als een groep met weinig genuanceerde interesses voor de dag zijn gekomen. Lage scores vinden we weer bij de beide vrouwelijke beroepen en voorts bij een aantal niet-technische beroepen.

Schaal VB die bij de constructie veel moeilijkheden had gegeven, in verband met de geringe homogeniteit van de items, blijkt toch zeer specifiek te appelleren aan de belangstelling van de beroepsbeoefenaars, die met de voeding te maken hebben. Alle betreffende beroepen bevinden zich aan de top. In negatieve zin laat zich hier weinig zeggen, omdat vele der onderzochte beroepen geen verband houden met de voedselbereiding.

Voor schaal AA geldt m.m. hetzelfde. De tuinman en bosopzichter bezetten de beide eerste plaatsen, gevolgd door enkele beroepsbeoefenaars, die agrarische producten verwerken.

In schaal HA is de eerste plaats, zoals te verwachten, ingeruimd voor de commerciële employés. Onverwacht zijn, naar Nederlands inzicht, de hoge plaatsen van de beide daarop volgende ambtenaarcategorieën. IRLE zegt hiervan echter, dat dit in overeenstemming is met de doelstellingen van de moderne ambtenarenselectie. Mogelijk speelt hier het cultuurverschil een rol. Bij de daaropvolgende beroepen is de hoogte van de score,

uitgedrukt in percentielen, reeds niet meer opvallend hoog, zodat hun positie geen interpretatie behoeft. Van de laagstgeplaatste beroepen kan men gemakkelijk inzien, dat commercieel inzicht hier over het algemeen geen rol speelt. Een uitzondering vormt het beroep van bosopzichter, waarin enig commercieel inzicht noodzakelijk is. Het ligt echter voor de hand, hier te veronderstellen, dat de betrokken jongelui hun beroep op niet-commerciële motieven kiezen en pas in de loop van hun carrière met het handelsaspect in aanraking komen.

De beroepen met een administratieve inslag vinden in de top van schaal AD hun plaats. De drukker bevindt zich ook hier op een punt, waar men eerder de letterzetter verwacht aan te treffen. Opvallend is ook de hoge rangplaats van de commercieel employé. De beroepen met een lage gemiddelde score zijn inderdaad specifiek niet-administratief, weer met uitzondering van de bosopzichter, die als neventaak enige administratie verricht. Dezelfde overweging als bij schaal HA zou ook hier kunnen gelden. In beide gevallen blijken bosopzichter en tuinman op dezelfde wijze te reageren, blijkbaar is de afkeer in de agrarische sector, zowel van administratief als van commercieel werk opvallend groot, hetgeen trouwens ook zijn weerspiegeling vindt in de hoogte van de negatieve correlatie tussen de schaal AA en de schalen HA en AD.

Bij de interpretatie van de gegevens over de schalen HA en AD dient nog te worden bedacht, dat in Duitsland de „afstand” tussen commerciële en administratieve arbeid minder groot is dan bij ons. De leerling-commercieel employé begint als regel zijn opleiding met een periode van administratief werk. Dit verklaart wellicht ook de positieve correlatie, die volgens tabel 2.2 tussen de beide gebieden bestaat.

Op schaal LG worden hoge scores behaald door de beroepen met een sterk verbale inslag. Aan de negatieve zijde van de schaal vinden wij vooral de metaalbewerksberoepen en voorts een aantal andere, waarin traditioneel het verbale element eveneens een zeer geringe rol speelt.

De hoogste scores op schaal SO worden door de beide vrouwelijke beroepen behaald. Opnieuw is er aanleiding aan te nemen, dat dit vooral als een sexe-kenmerk moet worden geïnterpreteerd, dat hier zo opvallend tot uiting komt, omdat het onderzoek zeer overwegend op manlijke beroepen betrekking heeft. Overigens behaalt alleen de groep leerling-gemeente ambtenaren een noemenswaardig hoge score. IRLE vermeldt nog, dat een groep pedagogen, die in ander verband onderzocht werd, eveneens hoge scores op de SO-schaal behaalde. Zij overtroffen de beide vrouwelijke groepen uit dit onderzoek nog aanmerkelijk. Aan de lage zijde van de schaal vinden we vooral de metaalbewerks en een aantal andere beroepsbeoefenaars, die veelal solitair werken. De lage score van de kellners behoeft geen verwondering; het is een bekend feit, dat in dit beroep de sociale belangstelling dikwijls uiterst gering is en door de ietwat serviele attitude nog verder wordt onderdrukt.

Indirect kan deze studie worden opgevat als een validatie van de test. IRLE zegt, dat voor een eigenlijke validatie onafhankelijke gegevens over de typische belangstellingsstructuur van de betreffende beroepen ter beschikking zouden moeten staan. Dergelijke gegevens zijn echter nog niet beschikbaar, zodat een kwantitatieve validatie achterwege moet blijven. Op dit probleem zullen wij later, in het kader van onze eigen beschouwingen over het validiteitsprobleem terugkomen.

Uit de rangordetabellen voor de negen schalen, kan eveneens een berekening van de intercorrelaties tussen de schalen geschieden. Ditmaal uiteraard in de vorm van rangordecoëfficiënten. Een berekening van deze coëfficiënten (SPEARMAN's rho) gaf resultaten, die zeer sterk overeenstemmen met de coëfficiënten, die eerder bij de berekening van de product-moment-correlaties werden verkregen. Een berekening van de rangcorrelatie tussen deze beide verzamelingen van correlaties gaf een waarde van $+ .87$. IRLE meent, dat dit mag worden opgevat als een aanwijzing, dat de verzameling beroepen, die hij in zijn onderzoek heeft betrokken, reeds bij goede benadering een afspiegeling vormt van de verscheidenheid, die er, ten opzichte van de interessen, in de beroepenwereld bestaat.

Hiermee willen wij dit overzicht van de studie van IRLE beëindigen. In zijn hoofdwerk heeft hij ook nog in het kort de vraag naar de verwantschap tussen interessenstructuur en intelligentiestructuur aan de orde gesteld. Dit is echter geschied op een wijze, die voor onze studie van weinig betekenis is. Er zal hier daarom niet nader op ingegaan worden.

Samenvattend kunnen we zeggen, dat door IRLE reeds een zeer omvangrijk en waardevol werk, ten bate van de ontwikkeling van zijn test voor praktische doeleinden, is uitgevoerd. Hij heeft zich echter geconcentreerd op een groep personen, die reeds enkele jaren de lagere school heeft verlaten. De leeftijdsgroep, die voor ons van eminente praktische betekenis is, heeft hij niet onderzocht. In de Duitse situatie, waar de beslissingen in verband met school- en beroepskeuze in de meeste landen eerst op het tiende en vervolgens op het veertiende jaar worden genomen, is de problematiek van de twaalfjarige niet van bijzonder belang. Hier grijpt nu ons eigen onderzoek aan. Na een bespreking van de test zal hiervan in de volgende hoofdstukken verslag worden uitgebracht.

HOOFDSTUK III

DE BEROEPENINTERESSE TEST

3.1. *Beschrijving van de test*

De B I T (zie bijlage V) bestaat uit een dubbelzijdig bedrukt vel papier, waarop aan iedere zijde 100 rechthoeken van $22 \times 15 \text{ mm}^2$ voorkomen. Deze rechthoeken zijn horizontaal en verticaal in 10 rijen en 10 kolommen gerangschikt. Tussen de rijen en kolommen is telkens een ruimte van ca. 3.5 mm opengelaten. Er ontstaat daardoor een regelmatig patroon, dat zou kunnen worden opgevat als de plattegrond van een moderne stad, met rechthoekig kruisende straten en huizenblokken. In ieder blok is een item geplaatst, volgens het schema, dat reeds in paragraaf 2.3 is gegeven. Aan ieder kruispunt van „straten” grenzen dus vier blokken met vier items, die tot verschillende interesserichtingen behoren. De testinstructie — die later in detail zal worden gegeven — luidt, een wandeling te maken door de „straten”, volgens een route, die met pijlen is aangegeven. Deze route is zo gekozen, dat alle straten worden doorlopen. Eerst horizontaal, heen en weergaand, en vervolgens verticaal, op dezelfde wijze. De wandeling begint op het kruispunt links bovenaan, dat daartoe met het cijfer 1 is gekenmerkt. De te onderzoeken persoon wordt uitgenodigd, de meest sympathieke, c.q. minst antipathieke van de vier items, die aan dit kruispunt grenzen, te merken. Dit geschiedt, door een kruis te plaatsen in een vierkantje, dat zich in iedere rechthoek aangrenzend aan ieder kruispunt bevindt. De pijlen volgend gaat de proefpersoon dan naar rechts, passeert het eerstvolgend kruispunt, dat gekenmerkt is als nr. 73 en derhalve voorlopig wordt overgeslagen en komt daarna op kruispunt nr. 2, waar opnieuw een keuze moet worden gedaan. Deze gang van zaken herhaalt zich steeds. Telkens wordt één kruispunt overgeslagen, nadat een keuze is gedaan. Dit brengt mee, dat de items, die in een keuzeproces betrokken zijn geweest, niet bij de onmiddellijk volgende keuze opnieuw moeten worden beoordeeld. Het ligt voor de hand, dat anders gemakkelijk van de voorafgaande keuze een beïnvloeding zou kunnen uitgaan. Nadat de wandeling langs de horizontale straten voltooid is, volgt de verticale weg, waarop de te voren overgeslagen kruispunten worden bezocht. Ook nu wordt ieder tweede kruispunt overgeslagen, dit zijn degenen die reeds op de horizontale wegen zijn beoordeeld.

Nadat aldus één bladzijde van de test is voltooid, wordt dezelfde gang van zaken op de achterzijde van het testformulier herhaald. Wanneer

de test in de zesde klasse van een lagere school klassikaal wordt afgenomen, is hiermee gemiddeld ongeveer één uur gemoeid.

3.2. *De vertaling van de B I T*

In ieder geval, waarin een test wordt vertaald, moet de vertaling aan bepaalde voorwaarden voldoen, die ten dele van algemene aard zijn en voor een ander deel samenhangen met de specifieke aard van de test. De algemene eis is uiteraard, dat de vertaling naar vorm en inhoud zo weinig mogelijk van het origineel afwijkt en liefst ook van dezelfde moeilijkheidsgraad is. In dit speciale geval komt daarbij in de eerste plaats de eis, dat de omvang van de vertaling zodanig moet zijn, dat iedere item in één blok van het testformulier kan worden afgedrukt. Dit brengt bijzondere moeilijkheden mee. De Duitse taal heeft, anders dan het Nederlands, het kenmerk dat samengestelde woorden gemiddeld op veel plastischer wijze kunnen worden gebruikt. Van een Duits samengesteld woord, dat ook aan de weinig vaktechnisch geschoolde lezer onmiddellijk aanspreekt, bestaat een Nederlands equivalent óf in het geheel niet, óf alleen als vakterm. Als voorbeeld noemen we „*Ferngasleitungen legen*”. Wellicht bestaat het woord „afstandsgasleiding” of iets soortgelijks als vakterm, maar dit is geen woord, dat men in een interesstest kan gebruiken. Wil men het begrip aan de niet-vaktechnisch geschoolde lezer duidelijk maken, dan moet men het omschrijven. In dat geval komt men echter in conflict met de eis van een beknopte formulering. De oplossing, die in dit geval is gekozen, is, een nauw verwant begrip te gebruiken: „Aardgasleidingen leggen”. Alleen wanneer het samengestelde woord betrekkelijk weinig „Aufforderungscharakter” heeft, kan men eventueel een meer technische vertaling handhaven. B.v. „*Kunststoffe auf Säurebeständigkeit prüfen*” kan worden vertaald met: „Plastics onderzoeken op zuurbestendigheid”. Het appèl van dit item gaat onmiskenbaar van de beide eerste woorden uit, de zuurbestendigheid is niet zeer essentieel voor de proefpersoon met natuurwetenschappelijke interesse.

Problemen ontstaan ook bij situaties, die internationaal niet geheel equivalent zijn. De zin „*Mit Radfahrern Unfallversicherungen abschliessen*” kan men wel zonder meer in het Nederlands vertalen, maar ieder kind hier te lande weet, dat onze fietsers hun leven onverzekerd plegen te wagen. Er is veel kans, dat in een dergelijk geval het appèl van het item hier anders zal zijn dan in een land, waar fietsers relatief veel zeldzamer verschijningen zijn. In dit geval is als vertaling gekozen: „Motorfietsers tegen ongevallen verzekeren”, waarbij dus tevens het samengestelde woord „ongevallenverzekering” wordt vermeden.

Als algemeen principe heeft bij de vertaling gegolden, dat deze zo letterlijk mogelijk de Duitse tekst moest volgen, echter met het voorbehoud van kernachtigheid, adaequaatheid in de Nederlandse situatie en begripelijkheid op het niveau van de gemiddelde twaalfjarige. In het volgende

overzicht worden naast elkaar de Nederlandse en de Duitse tekst weergegeven. De items zijn per interessegebied gegroepeerd. De nummering is geschied, door beginnend links bovenaan in de opeenvolgende kolommen de items van boven naar beneden te nummeren. De laatste item in iedere kolom, die identiek is aan de eerste, is niet genummerd; evenmin zijn de items van de laatste kolom, die identiek is met de eerste, genummerd. De laatste item van de eerste bladzijde draagt nummer 81, de nummering op de tweede bladzijde is op analoge wijze geschied, beginnende met nummer 82.

Technische handarbeid (TH)

- | | |
|--|--|
| 1. Bovenleidingen voor elektrische spoorwegen aanleggen. | Stromleitungen für elektrische Bahnen legen. |
| 4. Autolampen monteren. | Autoscheinwerfer einbauen. |
| 7. In nieuwe huizen centrale verwarming aanleggen. | Warmwasserheizungen in Neubauten anlegen. |
| 30. Brugpijlers uit beton storten. | Brückenpfeiler aus Beton giessen. |
| 33. Aan een draaibank werken. | An einer Drehbank arbeiten. |
| 36. Gereedschappen smeden. | Werkzeuge schmieden. |
| 57. Schrijfmachines monteren. | Schreibmaschinen zusammensetzen. |
| 60. Aardgasleidingen leggen. | Ferngasleitungen legen. |
| 63. Stoelen vervaardigen. | Stühle schreinern. |
| 82. Nieuwe beeldbuizen in televisietoestellen monteren. | Neue Bildröhren in Fernsehgeräte einsetzen. |
| 85. Lipssloten in deuren monteren. | Sicherheitsschlösser in Türen einbauen. |
| 88. In een rijwielfabriek frames lassen. | In Fahrradfabrik Rahmen schweissen. |
| 111. Scheepswanden klinken. | Schiffswände vernieten. |
| 114. Dieselmotoren repareren. | Dieselmotoren reparieren. |
| 117. In een nieuwe fabriek de machines monteren. | Maschinen in neuen Fabriken montieren. |
| 138. Krukassen voor auto's draaien. | Kurbelwellen für Autos drehen. |
| 141. Balken voor de huizenbouw op maat zagen. | Balken für Hausbauten zuschneiden. |
| 144. Transportkranen onderhouden. | Maschinenkräne instandhalten. |

Ambachtelijke vormgeving (AV)

- | | |
|--|--|
| 2. Reclamebiljetten voor een parfum ontwerpen. | Reklameplakate für ein Parfüm entwerfen. |
| 5. Glas-in-lood-ramen maken voor kerken. | Kirchenfenster farbig verglasen. |
| 8. Mooie etalages maken. | Schaufenster dekorieren. |
| 29. Felicitatiekaarten tekenen. | Glückwunschkarten zeichnen. |
| 32. Gouden armbanden smeden. | Armbänder aus Gold schmieden. |
| 35. Landkaarten tekenen. | Landkarten zeichnen. |

- | | |
|--|-----------------------------------|
| 55. Tabakspijpen uit hout snijden. | Tabakpfeifen aus Holz schnitzen. |
| 58. Porselein beschilderen. | Porzellan bemalen. |
| 61. Salonmeubels ontwerpen. | Pläne für Sitzmöbel entwerfen. |
| 83. Nieuwe kledingmodellen ontwerpen . | Neue Kleidermodelle ausdenken. |
| 86. Landschappen fotograferen. | Landschaften fotografieren. |
| 89. Bloemenvazen uit klei bakken. | Blumenvasen aus Ton brennen. |
| 110. Balcostuums ontwerpen. | Für einen Ball Kostüme entwerfen. |
| 113. Een danszaal versieren. | Einen Tanzsaal schmücken. |
| 116. Behangselpatronen ontwerpen. | Tapetenmuster entwerfen. |
| 136. Fraaie meubels vervaardigen. | Elegante Möbel tischlern. |
| 139. Landschappen schilderen. | Landschaftsbilder malen. |
| 142. Salonmeubelen ontwerpen. | Pläne für Sitzmöbel entwerfen. |

Techniek en Natuurwetenschap (TN)

- | | |
|---|---|
| 3. Plastics onderzoeken op zuurbestendigheid. | Kunststoffe auf Säurebeständigkeit prüfen. |
| 6. De sterkte van ijzeren draagbalken berekenen. | Tragfähigkeit von Eisenträgern berechnen. |
| 9. Het weer van verschillende steden vergelijken. | Wetter verschiedener Grosstädte vergleichen. |
| 28. Het traject van een nieuwe spoorwegtunnel bepalen. | Verlauf für einen Eisenbahntunnel festlegen. |
| 31. In een electrotechnisch laboratorium werken. | In elektrotechnischem Labor arbeiten. |
| 34. De haarden van aardbevingen bepalen. | Die Herde von Erdbeben bestimmen. |
| 56. Onderzoeken doen naar nieuwe steenkoolgebieden. | Nach neuen Kohlevorkommen forschen. |
| 59. Landmetingen voor een ondergrondse spoorweg verrichten. | Erdvermessungen für Untergrundbahnen durchführen. |
| 62. De remkracht van nieuwe remmen berekenen. | Bremswirkung neuer Bremsen berechnen. |
| 84. Erfelijkheidsproeven op planten doen. | Vererbungsexperimente an Pflanzen durchführen. |
| 87. Plantencellen onder een microscoop onderzoeken. | Unterm Mikroskop Pflanzenzellen untersuchen. |
| 90. Standplaatsen voor televisiemasten bepalen. | Standorte für Fernsehtürme festlegen. |
| 109. De electriche spanning van de bliksem meten. | Elektrische Spannungen von Blitze messen. |
| 112. Gasturbine's voor auto's construeren. | Turbinenantriebe für Autos konstruieren. |
| 115. Het prestatievermogen van motoren verbeteren. | Leistungsfähigkeit von Motoren verbessern. |
| 137. De nauwkeurigheid van thermometers testen. | Genauigkeit von Thermometern untersuchen. |

- | | |
|---|-------------------------------------|
| 140. Nieuwe toepassingen voor aardgas zoeken. | Verwendbarkeit von Erdgasen prüfen. |
| 143. Werken op een meteorologisch instituut. | An einer Wetterstation mitarbeiten. |

Voedselbereiding (VB)

- | | |
|---|--|
| 10. Toezicht houden op de vak-kundige bewaring van vlees in koelhuizen. | Sachgemässe Lagerung von Fleisch im Kühlhaus überwachen. |
| 13. Verfrissende vruchtensla maken uit zuidvruchten. | Erfrischende Obstsalate aus Südfrüchte zubereiten. |
| 16. Een wijnhandel leiden. | Eine Weinkellerei leiten. |
| 38. In een zuivelfabriek melk bacteriologisch onderzoeken. | In Molkerei Bakterienfreiheit der Milch überwachen. |
| 41. Prima koffiesoorten mengen. | Erlesene Kaffeesorten mischen. |
| 44. De keuken van een groot hotel leiden. | Küche eines grossen Hotels führen. |
| 65. De graanzolders van een maalderij beheren. | Getreidespeicher einer Mühle beaufsichtigen. |
| 68. Bier brouwen. | Als Brauereimeister Malzbieren brauen. |
| 71. Feestgerechten voor een bruiloftspartij bereiden. | Festliche Gerichte für Hochzeitsfeiern bereiten. |
| 91. Indische specerijen voor de invoer uitzoeken. | Indische Gewürze zur Einfuhr auswählen. |
| 94. Verfrissende vruchtensla uit zuidvruchten maken. | Erfrischende Obstsalate aus Südfrüchte zubereiten. |
| 97. Ossen aan het spit braden. | Ochsen am Spiess braten. |
| 119. Prima koffiesoorten mengen. | Erlesene Kaffeesorten mischen. |
| 122. Slachtvee op ziekte keuren. | Schlachtvieh auf Gesundheitszustand untersuchen. |
| 125. Een banketbakkerij leiden. | Eine Konditorei leiten. |
| 146. In een proefkeuken diëtrecepten proberen. | In Versuchsküche Diätrezepte erproben. |
| 149. Een snack-bar oprichten. | Eine Milchbar einrichten. |
| 152. Scheikundige onderzoeken doen over levensmiddelen. | Nahrungsmittelchemische Forschungen durchführen. |

Agrarische arbeid (AA)

- | | |
|---|--------------------------------|
| 12. Eieren uitbroeden in een broed-machine. | Küken in Brutöfen ausbrüten. |
| 15. Rozen stekken. | Rosenstöcke schneiden. |
| 18. Beukenbossen uitdunnen. | Buchenwälder auflichten. |
| 39. Bossen aanplanten. | Nutzholzforste anpflanzen. |
| 42. Kasbloemen kweken. | Blumen im Gewächshaus züchten. |
| 45. Karpervijvers aanleggen. | Karpfenteiche anlegen. |
| 64. Dieren in een dierentuin verzorgen. | Tiere in einem Zoo pflegen. |

67. Bloembedden aanleggen.	Tulpenbeete anlegen.
70. Op de haringvangst gaan.	Auf Heringsfang fahren.
93. Rashonden fokken.	Rassehunde züchten.
96. Zwarte tulpen kweken.	Schwarze Tulpen züchten.
99. Een vruchtboomgaard aanleggen.	Eine Obstplantage anlegen.
120. Forellen vangen.	Forellen fischen.
123. Dennenbossen in goede staat houden.	Tannenwälder aufforsten.
126. Orchideeën kweken.	Im Gewächshaus Orchideen züchten.
145. Bomen uitzoeken voor het omhakken.	Bäume zum Fällen aussuchen.
148. Vruchtbomen bevrijden van schadelijke insecten.	Obstbäume von Schädlingen befreien.
151. Een bijenteelt beginnen.	Eine Bienenzucht anlegen.

Handel (HA)

11. De kaartverkoop in een circus organiseren.	Kartenverkauf für Zirkus organisieren.
14. Auto's verkopen.	Verkauf von Personenwagen steigern.
17. De klanten van een bank adviseren bij beleggingen.	Bankkunden über Vermögensanlage beraten.
37. Motorfietsers tegen ongevallen verzekeren.	Mit Radfahrern Unfallversicherungen abschliessen.
40. Nieuwe technische vindingen bij klanten introduceren.	Technische Neuheiten bei Kunden einführen.
43. Verkopen leiden.	Versteigerungen leiten.
66. Verkoopbrieven schrijven.	Werbepriefe an Kunden schreiben.
69. Koffie inkopen voor een koffiebrandrij.	Kaffee für Röstereien einkaufen.
72. Koelkasten verkopen.	Käufer für Kühlschränke gewinnen.
92. Krantenreclame voor tandpasta maken.	Zeitungswerbung für Zahnpasta durchführen.
95. Handelen in goud en sieraden.	Mit Schmuck und Gold handeln.
98. Met boeren brandverzekeringen afsluiten.	Mit Bauern Feuerversicherungen abschliessen.
118. Goede verkoopmanieren leren aan verkopers.	Verkäufer zu gutem Kundendienst anleiten.
121. Klanten raad geven bij de aankoop van meubelen.	Kunden beim Möbelkauf beraten.
124. Artisten verzekeren tegen ongevallen.	Artisten gegen Unfall versichern.
147. Deelnemers werven voor een vacantiereis naar Italië.	Für Ferienreisen nach Italien werben.
150. De klanten van een bank bedienen.	Publikum in einer Bank bedienen.
153. Werken als verkoper voor een autohandel.	Als Vertreter für Autofirma arbeiten.

Administratief werk (AD)

- | | |
|---|--|
| 19. Loon- en salarisberekeningen maken. | Gehalts- und Lohnabrechnungen durchführen. |
| 22. Rapporten maken over de omzet van goederen. | Berichte über Warenumsatz schreiben. |
| 25. Notaris-acten opstellen. | Verwaltungsakten bearbeiten. |
| 48. Aanvragen voor bouwvergunningen behandelen. | Anträge auf Baugenehmigung bearbeiten. |
| 51. De goederenomzet van een zaak berekenen. | Warenumsatz eines Geschäftes berechnen. |
| 54. De goederenverzending in een groothandel voorbereiden. | Warenversand im Grosshandel vorbereiten. |
| 75. In een magazijn voorraadlijsten maken. | In Warenlager Bestandslisten anfertigen. |
| 78. Als belastingadviseur voor bedrijven werken. | Als Steuerberater für Industriefirma arbeiten. |
| 81. Loonbelasting berekenen. | Lohnsteuerberechnungen durchführen. |
| 100. De goederenexpeditie van een papierhandel leiden. | Auslieferungslager im Papierhandel leiten. |
| 103. Werken met een rekenmachine. | An einer Rechenmaschine arbeiten. |
| 106. De klantenadministratie van een groothandel bijhouden. | Kundenkartei im Grosshandel bearbeiten. |
| 129. Op een belastinginspectie de belastingaanslagen regelen. | In einem Finanzamt Steuereinnahmen berechnen. |
| 132. De toename van de tegoeden bij spaarbanken onderzoeken. | Zunahme der Sparguthaben untersuchen. |
| 135. De productiekosten van een fabriek berekenen. | In einer Fabrik Produktionskosten berechnen. |
| 156. In een verzendhuis de postzegelkas beheren. | In Versandgeschäft Portokasse führen. |
| 159. De goederenontvangst en -verzending bij een firma controleren. | Waren-Ein- und Ausgänge einer Firma kontrollieren. |
| 162. Loonafrekeningen controleren. | Lohnabrechnungen nachprüfen. |

Literaire en geesteswetenschappelijke arbeid (LG)

- | | |
|--|---|
| 21. Geschiedenisles geven. | Geschichtsunterricht erteilen. |
| 24. Korte verhalen schrijven. | Kurzgeschichten schreiben. |
| 27. Buitenlandse romans vertalen. | Ausländische Romane übersetzen. |
| 46. Radiohoorspelen schrijven. | Hörspiele für Rundfunk schreiben. |
| 49. Amateur-toneelvoorstellingen regisseren. | Laientheaterspiele leiten. |
| 52. Aan een volksuniversiteit cursussen geven over geschiedenis. | Volkshochschulkurse über Geschichte halten. |
| 74. Verhalen schrijven voor geïllustreerde tijdschriften. | Aufsätze für Illustrierte schreiben. |

- | | |
|---|--|
| 77. Waardevolle nieuwe gedichten bundelen. | Wertvolle neue Gedichte sammeln. |
| 80. Sagen en sprookjes verzamelen. | Sagen und Märchen sammeln. |
| 102. Korte verhalen schrijven voor kranten. | Kurzgeschichten für Zeitungen schreiben. |
| 105. Reportages schrijven voor dagbladen. | Berichte für Tageszeitungen schreiben. |
| 108. Draaiboeken voor films schrijven. | Drehbücher für Filme schreiben. |
| 127. De oorzaken van de zevenjarige oorlog naspeuren. | Ursachen des Siebenjährigen Krieges untersuchen. |
| 130. Humoristische verzen maken. | Humoristische Verse verfassen. |
| 133. Krantenverslagen over nieuwe films schrijven. | Zeitungsberichte über neue Filme verfassen. |
| 155. Biografieën van beroemde mensen verzamelen. | Lebensberichte berühmter Männer sammeln. |
| 158. Een geschiedenisboek over de roofriddertijd schrijven. | Geschichtsbuch über Raubritterzeit schreiben. |
| 161. Toneelcritieken schrijven. | Theaterkritiken schreiben. |

Sociaal werk en opvoeding (SO)

- | | |
|---|---|
| 20. Kinderverlamningspatiënten verplegen. | An Kinderlähmung Erkrankte pflegen. |
| 23. Moeilijk-opvoedbare kinderen verzorgen. | Schwererziehbare Kinder betreuen. |
| 26. Arme mensen ondersteunen. | Unterstützungen an Bedürftige verteilen. |
| 47. Blindgeworden chauffeurs een nieuw beroep leren. | Erblindete Kraftfahrer neue Berufe lehren. |
| 50. Gevangenen helpen bij hun terugkeer in de maatschappij. | Gefängnisentlassenen ins Leben zurückhelfen. |
| 53. Loongeschillen beslechten. | Lohnstreitigkeiten schlichten. |
| 73. Jeugdige gevangenen opvoeden. | Jugendliche Gefängnisinsassen erziehen. |
| 76. Kleuterscholen inrichten. | Kindergärten einrichten. |
| 79. Landlopers opvoeden tot arbeidzaamheid. | Landstreicher zur Arbeitsamkeit erziehen. |
| 101. Ongevalspatiënten verplegen. | Unfallsverletzte pflegen. |
| 104. Grote gezinnen een passende woning bezorgen. | Wohnungen für kinderreiche Familien besorgen. |
| 107. Jeugdige misdadigers op het goede pad brengen. | Kriminelle Jugendliche zum Guten führen. |
| 128. Pleegouders zoeken voor weeskinderen. | Für Waisenkinder Pflegeeltern suchen. |
| 131. Toezicht houden op kinderen van werkende moeders. | Kinder berufstätiger Mütter beaufsichtigen. |

- | | |
|---|---|
| 134. Vluchtelingen helpen een nieuwe woonplaats te vinden. | Heimatlose Menschen sesshaft machen. |
| 154. Geestelijk-ongelukkige mensen helpen. | Seelisch unglücklichen Menschen helfen. |
| 157. Gebrekkige mensen rondrijden. | Verkrüppelte Menschen ausfahren. |
| 160. Voor zieken een plaats in een herstellingsoord zoeken. | Erholungsaufenthalte für Kranke vermitteln. |

Bij nauwkeurige bestudering van bovenstaande lijst zal blijken, dat tussen origineel en vertaling in een aantal gevallen verschillen bestaan, die niet voortvloeien uit de eerstgenoemde principes. Aan de hand van een enkel voorbeeld willen wij toelichten, welke overwegingen daarbij een rol kunnen hebben gespeeld. Nr. 123 luidt in het origineel: „Tannenwälder aufforsten”. Daargelaten of hier met „Tanne” een spar of een den wordt bedoeld, zal het eerste woord in het Nederlands in ieder geval met „dennenbos(sen)” moeten worden vertaald. „Sparrenbos”, „dennenwoud” zijn te weinig gangbare woorden, in verhouding tot het gangbare „dennenbos”. Het woord „aufforsten” zou letterlijk moeten worden vertaald met „bebossen”. Dit zou echter een lelijke samenklank met het eerste woord geven, terwijl bovendien de betekenis van het woord voor de gemiddelde lezer onduidelijk is. Een alternatief van nagenoeg gelijke betekenis zou zijn: „de kwaliteit van dennenbossen verbeteren”, maar deze vertaling werd te moeilijk geacht. „Dennenbossen verbeteren” vermijdt het moeilijke woord „kwaliteit”, maar is zo beknopt, dat de betekenis opnieuw obscuur wordt. Een andere mogelijke vertaling zou zijn: „Dennenbossen in betere staat brengen”, maar ook dit is nog een rijkelijk moeilijke formulering. Daarom is tenslotte de voorkeur gegeven aan een vertaling, die naar inhoud iets van het origineel afwijkt, maar daar toch, wat het appèl aan de belangstelling betreft, vermoedelijk nauw bij zal aansluiten: „Dennenbossen in goede staat houden”.

Het zou te ver voeren, op deze wijze alle verschillen tussen origineel en vertaling toe te lichten. Van situatie tot situatie zijn de overwegingen ook verschillend geweest; zo zijn b.v. in enkele gevallen, waar het activiteiten met sterk intellectuele inslag betrof, vertalingen gebruikt, die hoger eisen stellen aan het begripsvermogen, dan elders het geval is. Wij zijn ons er van bewust, dat in verschillende gevallen gelijkwaardige alternatieven aanwezig zijn of zelfs vertalingen, die achteraf gezien de voorkeur hadden verdiend. Hoezeer ook naar een zorgvuldige vertaling wordt gestreefd, dergelijke verschijnselen zullen zich blijven voordoen. Zolang het betreffende uitzonderingsgevallen betreft, zal de eventuele storende invloed hiervan door de ijking van de test praktisch worden opgeheven.

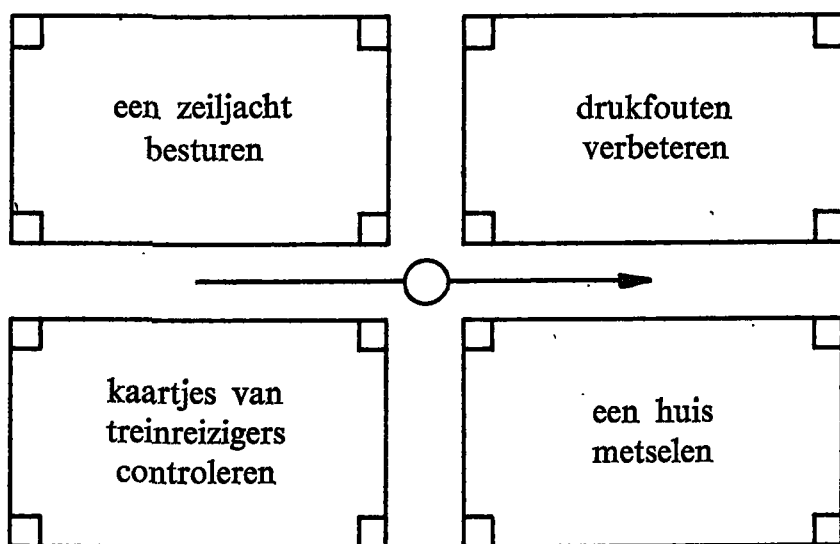
3.3. Handleiding voor de B I T

Ter verduidelijking van de hierna volgende handleiding, wordt verwezen naar het als bijlage opgenomen exemplaar van het testformulier. Uitgegaan

wordt van de gedachte, dat de test klassikaal wordt afgenomen, bij 12- à 14-jarige leerlingen. In andere situaties moeten vanzelfsprekend adequate wijzigingen in de instructie worden aangebracht. Verondersteld wordt voorts, dat de test wordt afgenomen in het kader van een meer omvattend psychologisch onderzoek, zodat het onderzoek met de BIT geen speciale introductie behoeft. Wordt de test afzonderlijk gegeven, dan blijken in de praktijk een zeer korte inleiding over de bedoeling van het onderzoek en enkele algemene aanwijzingen over de gang van zaken voldoende, als voorbereiding op de testinstructie.

Nadat de kinderen allen een testformulier voor zich hebben, luidt de instructie: „Jullie ziet op het blad, dat ik je gegeven heb, een groot aantal vakjes. Die staan op rijen, zoals de huizen in straten, die elkaar kruisen. In ieder vakje staat een bezigheid en op ieder kruispunt een nummer en een pijl. We gaan door deze straten wandelen en beginnen op het punt waar een 1 staat, in de linker bovenhoek. Op dat eerste kruispunt zie je vier vakjes, met vier verschillende bezigheden. Eén van de vier zal je het best bevallen, die zou je het liefst willen doen. Zet in dat vakje een kruisje, in het hokje dat aan het kruispunt grenst”.

De proefleider dient dit toe te lichten aan een voorbeeld, dat op het bord wordt getekend (figuur 3.1).



figuur 3.1

847

De proefleider demonstreert, hoe in één van de hokjes, *grenzend aan het kruispunt*, een kruisje wordt gezet. Hij vervolgt met:

„Als je bij het eerste kruispunt je keuze hebt gedaan, loop je in de rich-

ting van de pijl verder, tot je het kruispunt tegenkomt, waarbij nummer 2 staat. Daar kun je weer op dezelfde manier een keuze doen. Je zet weer een kruisje in het vakje met de bezigheid, die je het liefst zou willen doen. En dan ga je weer door, in de richting van de pijl, naar nummer 3 en zo steeds verder. Je zult zien, dat je telkens één kruispunt moet overslaan, om weer bij het goede nummer te komen. Als je aan het einde van de straat bent gekomen, ga je in de richting van de pijl ombuigen en dan kom je vanzelf weer op het volgende kruispunt.

Het kan natuurlijk gebeuren, dat geen van de vier bezigheden je erg leuk lijkt. Dan zal er toch wel één zijn, die je nog het liefste zou willen doen, die kies je dan. Het kan ook voorkomen, dat alle vier dingen aantrekkelijk zijn. In dat geval kies je degene, die het allerprettigst is. Op ieder kruispunt moet je dus één en niet meer, maar ook niet minder hokjes aankruisen.

Je zult zien, dat je in de loop van de wandeling iedere bezigheid viermaal tegenkomt. Je mag hem dus ook viermaal kiezen als hij je erg leuk toelijkt, maar dat hoeft natuurlijk niet. Telkens als je hem kiest, zet je een kruisje in het hokje, dat grenst aan het kruispunt waar je juist gekomen bent. Er zijn geen goede of foute antwoorden! Je kunt dus kiezen wat je het leukst vindt! Aarzel niet te lang met je beslissing; kies maar wat je op het eerste gezicht het meest aantrekkelijk voorkomt. Stel je bij het invullen maar voor, dat alle bezigheden even goed betaald worden en dat zij ook allemaal evenveel aanzien bij de mensen genieten.

Heeft iemand nog iets te vragen? Als er iets nog niet begrepen is, zal ik het graag nog eens uitleggen. . . . Alles goed begrepen? Begin dan met de voorkant. Als je bij kruispunt 81 bent gekomen, keer dan het blad om, dan zie je aan de achterkant nog een soortgelijke opgave. Denk er om, dat je geen kruispunt mag overslaan!"

Voorafgaand aan de instructie worden zo nodig de personalia aan de voorzijde van het formulier ingevuld. Nadat de groep is begonnen, maakt de proefleider een ronde om zich ervan te overtuigen, dat ieder de opdracht goed heeft begrepen. Zonodig wordt een nadere instructie gegeven. De proefleider moet er vooral op letten, of er kinderen zijn, die teveel hokjes invullen of die ten onrechte kruispunten overslaan. Sporadisch komt het voor, dat een kind de kruisjes niet in de daarvoor bestemde hokjes zet, maar b.v. midden in de vakjes. Ook dit is ongewenst; het kind moet dan worden gewezen op de juiste wijze van invullen. Indien een leerling reeds zoveel fouten heeft gemaakt, dat hierdoor bij de scoring ernstige moeilijkheden zouden ontstaan, is het gewenst dit kind een nieuw formulier te geven en het opnieuw te laten beginnen. Afkijken komt slechts zelden voor, uiteraard is het ongewenst.

Ook in het verdere verloop van het onderzoek doet de proefleider geregeld een ronde door de klas, waarbij hij speciaal op fouten tegen de testinstructie let. Zeer traag werkende kinderen worden eventueel aangespoord tot een hoger tempo, b.v.: „Niet te lang aarzelen; kies datgene,

wat je op het eerste gezicht het leukste lijkt!” In de meeste stedelijke gebieden zal na ongeveer een half uur vrijwel iedereen de eerste helft van de test voltooid moeten hebben. In sommige plattelandsdistricten kan het tempo gemiddeld wat lager liggen. Wanneer het kind gereed is, dient de proefleider zich door globale inspectie ervan te overtuigen, dat beide zijden van het formulier zijn bewerkt en dat geen kruispunten zijn overgeslagen. Het is gewenst, de tijd die het kind voor de test nodig heeft gehad, op het formulier te noteren.

Het nazien van de testformulieren kan het beste geschieden met behulp van schablonen, die men eventueel zelf van karton kan vervaardigen. Iedere „stem”, die op een item is uitgebracht, wordt voor één punt gerekend, de punten van alle items, die tot één schaal behoren, worden gesommeerd. De interpretatie van de scores geschiedt met behulp van ijkingsgegevens. Deze zijn voor gemiddeld ongeveer 16-jarigen door IRLE gepubliceerd (⁶⁷), terwijl zij voor kinderen omstreeks het einde van de lagere school in de volgende hoofdstukken van dit boek, in het bijzonder hoofdstuk VI, zijn aan te treffen.

HOOFDSTUK IV

NIVEAUVerschillen in de belangstelling

4.1. *Probleemstelling*

Het niveaubegrip speelt in de beroepenkunde reeds geruime tijd een belangrijke rol. Bij pogingen om te komen tot een beroepenclassificatie, tracht men veelal mede gebruik te maken van een verticaal ordeningsprincipe. Het ene beroep eist meer intelligentie dan het andere, heeft een langere opleidingsduur, vraagt meer financiële en morele steun van de omgeving om een gemiddelde kans van welslagen te hebben, geeft meer maatschappelijk aanzien, geeft een groter gemiddeld inkomen, etc. Deze verschillende gezichtspunten zijn niet onderling onafhankelijk ⁽⁹⁶⁾, er bestaat integendeel een vrij hoge correlatie. Vandaar dat een verticale ordening naar verschillende der genoemde principes wel in details divergenties vertoont ⁽¹⁴³⁾, maar toch in grote lijnen overeenkomstige resultaten oplevert.

Men kan zich nu afvragen, of ook op grond van de belangstelling een soortgelijke ordening mogelijk is. Indien dit het geval is, zou men een niveauschaal kunnen construeren, op grond waarvan reeds bij de voorlichting bij beroepskeuze een oriëntatie kan worden gegeven, betreffende het niveau, waarop gezien de interessen, het te kiezen beroep ongeveer zou moeten liggen. Pogingen, tot een dergelijke niveauschaal te komen, zijn het eerste in Amerika ondernomen: STRONG heeft in het kader van zijn test een aantal „occupational level” schalen geconstrueerd, waarvan er één definitief als bruikbaar is aanvaard. Bij de constructie van deze schaal is uitgegaan van de belangstelling van ongeschoolde arbeiders enerzijds en van een aantal groepen van academici en andere beoefenaars van hogere beroepen anderzijds. Gebleken is, dat gemiddeld de beoefenaars van beroepen, die in de classificatie meer een middenpositie innemen, ook op de niveauschaal scores behalen, die tussen de beide genoemde uitersten liggen. Wel is opgevallen ⁽¹²⁴⁾, dat de spreiding van de „occupational level” schaal binnen ieder beroep aanzienlijk groter is dan bij de specifieke beroepsschalen. STRONG wijst er op, dat in verband hiermee enige voorzichtigheid bij de individuele toepassing van de schaal gewenst is, maar staat toch niet afwijzend tegenover een dergelijk gebruik.

De Amerikaanse onderzoekingen hebben in hoofdzaak op volwassenen betrekking gehad. De jongste onderzochte groepen, waarover gegevens gerapporteerd zijn, bestaan uit 15-jarigen. Maar bij de kinderen, die het

lager onderwijs gaan verlaten, komen wij ook reeds met het probleem van de „occupational level” in aanraking. De keuze van de vorm van voortgezet onderwijs die men gaat bezoeken, is reeds van uitermate grote betekenis voor het niveau, dat men straks in het beroepsleven zal bereiken. Weliswaar blijft nog gedurende enkele jaren een overgang naar andere vormen van voortgezet onderwijs mogelijk en is tegenwoordig in verschillende vakrichtingen de mogelijkheid geopend, om ook zonder een langdurige algemeen vormende vooropleiding een middelbaar of academisch niveau te bereiken. Ongetwijfeld hebben ook allerlei persoonlijke kwaliteiten een grote invloed op het verloop van de carrière. Maar desondanks moet toch worden gesteld, dat de gemiddelde verwachting ten aanzien van het toekomstig beroep van een kind, dat een instelling voor v.h.m.o. gaat bezoeken, anders is dan b.v. die ten aanzien van een leerling, die naar een lagere technische school gaat.

De differentiatie aan het einde van de lagere school heeft daardoor een tweeledig aspect. Enerzijds is het een differentiatie naar niveau; algemeen wordt erkend, dat de eisen, die het v.h.m.o. stelt zwaarder zijn, dan die van andere vormen van voortgezet onderwijs. Traditioneel wordt zelfs zeer algemeen een verder gedifferentieerde rangorde aangehouden: v.h.m.o. — m.u.l.o. — vakonderwijs — v.g.l.o. (als eindonderwijs). Anderzijds is echter aan deze indeling ook een differentiatie naar richting inherent: de onderscheiding tussen meer theoretisch en meer praktisch, tussen handarbeid en hoofdarbeid, tussen overwegend symbolisch-abstract c.q. overwegend verbaal en overwegend concreet denken. Het is niet mogelijk, dit kwalitatief aspect scherp te scheiden van het niveau-aspect, zij zijn wezenlijk verbonden.

In dit hoofdstuk zullen wij ons desondanks concentreren op het niveau-aspect, met alle erkenning van het feit, dat daarmee impliciet ook aan kwalitatieve verschillen wordt geraakt. We gebruiken echter, om terminologische verwarring te voorkomen, het woord richting uitsluitend in verband met de attitude ten opzichte van de grote sectoren uit de beroepswereld, zoals deze in de negen belangstellingsschalen van IRLE zijn aangeduid. Het woord niveau wordt, eveneens specifiek betrokken op het feit, dat verschillende vormen van voortgezet onderwijs elkaar als het ware laagsgewijze overdekken, met name in die zin, dat zij aan velerlei aspecten van de begaafdheid en de werkhouding successievelijk hoger eisen stellen.

De vraag die wij dan te beantwoorden krijgen is, of er in de B I T verschillen in voorkeur op te merken zijn, die samenhangen met niveauverschillen. Dergelijke verschillen zouden eventueel kunnen dienen als basis voor de constructie van één of meer niveauschalen, in de geest van de „occupational level” schaal. Ten behoeve van het hier bedoelde onderzoek zullen wij drie niveau's onderscheiden:

1. *Lager voortgezet onderwijs* (L-niveau). Hiertoe rekenen we het lager technisch onderwijs, de nijverheidsscholen voor meisjes, het lager land-

en tuinbouwonderwijs en een aantal zeldzamer voorkomende typen van voorbereidend vakonderwijs. Bovendien wordt ook het v.g.l.o. tot deze categorie gerekend, hoofdzakelijk op grond van de overweging, dat een niet onbelangrijk gedeelte van de leerlingen van dit onderwijs later nog het vakonderwijs bezoekt, of in de praktijk, in het kader b.v. van een leerlingstelsel een vakopleiding krijgt.

2. *Uitgebreid lager onderwijs* (U-niveau). Hiertoe rekenen we het handelsdagonderwijs en het (m.)u.l.o. De in enkele steden voorkomende u.l.o.-scholen, die niet opleiden voor een erkend m.u.l.o.-diploma, worden door ons tot het lager voortgezet onderwijs gerekend.
3. *Middelbaar voortgezet onderwijs* (M-niveau). Dit omvat de overige vormen van voorbereidend hoger en middelbaar dagonderwijs.

Met deze indeling kan ieder kind, dat na de lagere school een vorm van voortgezet onderwijs gaat bezoeken, worden geclassificeerd. Nagegaan zal worden, of er in onze belangstellingstest items zijn, die tussen de niveau's discrimineren en voorts of het mogelijk zal zijn een schaal te construeren, die in het bijzonder in het kader van de differentiatie aan het einde van de lagere school een praktisch bruikbaar hulpmiddel zal blijken te zijn. Een dergelijk hulpmiddel kan van waarde zijn, als het er om gaat, voorlichting te geven aan ouders en kinderen bij de schoolkeuze.

Onze taak zal dus in de eerste plaats zijn, items op te sporen, die differentiërend werken. Op deze basis kan op tweeërlei wijze verder worden gewerkt. In de eerste plaats kan men trachten, met behulp van deze items een belangstellingsschaal te construeren, analoog aan de schalen, die reeds in de B I T worden gebruikt. Een tweede mogelijkheid is, te trachten, een schaal met gewogen items te maken, op het voetspoor van STRONG en tegenwoordig ook KUDER. Indien slechts één dergelijke schaal in een test wordt opgenomen, zijn de praktische consequenties nog niet zodanig bezwaarlijk, dat de toepassing hierop zou moeten afstuiten. In verband met het feit, dat de constructie van een dergelijke „gewogen schaal” op sterk afwijkende wijze geschiedt, is het gehele, daarop betrekking hebbende onderzoek afzonderlijk uitgevoerd. Het zal eerst in het laatste gedeelte van dit hoofdstuk aan de orde worden gesteld. De eerste paragrafen hebben betrekking op de problematiek van de constructie van een schaal van het traditionele type.

4.2. *Eerste experimentele opzet*

Teneinde items te vinden, die naar niveau differentiëren tussen kinderen omstreeks het einde van de lagere school, zullen we vergelijkend te werk moeten gaan. Vanzelfsprekend zal deze vergelijking betrekking moeten hebben op kinderen, bij wie het niveau van het voortgezet onderwijs reeds in voldoende mate vast staat. Dergelijke kinderen zullen groepsgewijs met behulp van de B I T moeten worden onderzocht. Daarna zal item voor

item moeten worden nagegaan, of er voorkeursverschillen tussen de proefgroepen, in het bijzonder tussen de extreme niveau's, bestaan.

De eerste vraag, die wij moeten beantwoorden, betreft het criterium voor de selectie van onze proefpersonen. We willen de test gebruiken in het kader van de differentiatie aan het einde van de lagere school; de proefpersonen dienen daarom bij voorkeur naar ontwikkelingsstadium ook ongeveer op dat punt te zijn gekomen. Anderzijds echter moet bij de proefpersonen het niveau van het voortgezet onderwijs, dat zij gaan volgen, reeds vast staan. Dit laatste is evenwel in de zesde klasse van de lagere school bij een aanzienlijk aantal leerlingen nog niet het geval. Ongetwijfeld kan men reeds in de loop van de zesde klas ten aanzien van vele leerlingen een uitspraak doen over de best-passende schoolkeuze voor het voortgezet onderwijs. De ervaring leert echter, dat er ook een vrij hoog percentage kinderen is, ten aanzien van wie nog geen uitspraak kan worden gedaan. Nog belangrijker in verband met ons onderzoek is wel, dat er een niet gering aantal gevallen is, waarin het oordeel zowel van het kind of zijn ouders, als van de klasse-onderwijzer, in de loop van het zesde leerjaar een wijziging ondergaat. Indien dergelijke gevallen in ons experimenteel materiaal zouden voorkomen, zou dit de uitkomsten van het onderzoek onzuiver maken. De correctie van fouten in de grondslag is wel mogelijk, maar is tijdrovend en kostbaar. Daarom werd besloten het onderzoeksmateriaal niet te verzamelen in de lagere school, maar bij leerlingen van de eerste klassen van scholen voor voortgezet onderwijs.

De vraag is dan echter, of deze leerlingen in hun reacties nog representatief zijn voor de interesses van de zesde klassers. Zij verschillen daarvan immers in twee opzichten. Gemiddeld zijn zij iets ouder dan de corresponderende groep in de zesde klasse en bovendien hebben zij de vormende en differentiërende invloed van hun nieuwe schoolomgeving ondergaan. In beide opzichten is het moeilijk, de invloed vooraf goed te taxeren. Validatie-onderzoek zal er achteraf wel inlichtingen over kunnen verstrekken, maar is eerst mogelijk, als het gehele werk van de testconstructie reeds is verricht.

Het is echter mogelijk de invloed van de beide verschilpunten bij voorbaat tot een minimum te reduceren, door de proefpersonen zeer kort na hun intrede in de nieuwe school te onderzoeken. Men kan dan met vrij grote zekerheid veronderstellen, dat de invloed van de leeftijdsfactor verwaarloosbaar klein zal worden en dat de invloed van de schoolkeuze op de belangstelling nog niet ver zal hebben doorgewerkt.

Bij voorbaat mag worden aangenomen, dat op ieder schooltype van het voortgezet onderwijs een aanzienlijk aantal leerlingen aanwezig zal zijn, die naar belangstelling sterk van het gemiddelde afwijken. Bovendien zijn er vooral in de aanvangsperiode vele leerlingen, die zich op het betreffende schooltype niet zullen kunnen handhaven. Daardoor zal er vanzelfsprekend een aanzienlijke variantie binnen de groepen zijn, die de variantie

tussen de groepen overdekt. Verschillen tussen de groepen zullen dus in het algemeen pas duidelijk aan het licht treden, bij een onderzoekmateriaal van aanzienlijke omvang. Hiertegenover kan worden gesteld, dat het onderzoek een praktisch doel beoogt, namelijk de constructie van een schaal, die als hulpmiddel bij de individuele differentiatie kan dienen. De significante verschillen in voorkeur moeten dus toch zo geprononceerd zijn, dat zij reeds bij een niet al te grote omvang van de proefgroepen statistisch kunnen worden geconstateerd. Deze beide tegengestelde overwegingen pleiten voor een materiaal van gemiddelde omvang.

In eerste instantie werd uitgegaan van de gedachte, dat iedere niveau-groep ongeveer 100 kinderen zou moeten omvatten. Aangezien met een klassikale test wordt gewerkt, zal het werkelijke aantal proefpersonen een benadering van dit getal zijn.

Ten aanzien van de leerlingen op middelbaar en uitgebreid lager niveau hebben wij ons inderdaad aan deze richtlijn gehouden. Bij het voortgezet lager onderwijs rees echter de moeilijkheid, dat hieronder een grote verscheidenheid van schooltypen valt. Een steekproef, waarin de verschillende vormen van voortgezet lager onderwijs enigszins representatief zijn vertegenwoordigd, is lastig te construeren. Daarom hebben wij ons beperkt tot de twee belangrijkste vormen, het nijverheidsonderwijs voor meisjes en het lager technisch onderwijs voor jongens. Dit heeft ongetwijfeld enige invloed op de uitkomsten van het onderzoek. Met name kan worden verwacht, dat de uitsluiting van leerlingen van het lager land- en tuinbouwonderwijs een vermindering van het aantal voorkeursreacties ten opzichte van de items uit de agrarische sector, zal meebrengen. Dit tekort moet worden erkend, hoewel de praktische betekenis waarschijnlijk niet zeer groot is. Uit het vervolg zal blijken, dat ook nu reeds ten aanzien van een aanzienlijk aantal agrarische items significante voorkeursverschillen ten gunste van de leerlingen van het voortgezet lager onderwijs optreden.

Binnen het lager technisch onderwijs moet nog weer onderscheid worden gemaakt tussen een groot aantal vakrichtingen. In dit opzicht waren wij gelukkig, de medewerking te verwerven van een school met een grote verscheidenheid van opleidingen. In ons materiaal zijn daardoor houtbewerkers, metaalbewerkers, electriciens, schilders, banketbakkers, boekdrukkers en zetters vertegenwoordigd. Mede door deze verscheidenheid is het aantal jongens, dat op lager technisch niveau werd onderzocht, ruim 150 geworden.

In verband met dit aanzienlijke aantal onderzochten zijn geen v.g.l.o.-scholen in het onderzoek opgenomen. Kwantitatief zal dit vermoedelijk geen grote invloed hebben gehad, gezien het in het algemeen lage begaafdheidsniveau van deze leerlingen en het welbekende feit, dat bij lage begaafdheid een belangstellingstest over het algemeen niet meer goed differentieert (^{83a}, pag. 3).

De proefpersonen voor het onderzoek waren leerlingen van scholen in

verschillende plaatsen in Noord- en Zuid-Holland. De numerieke samenstelling van de groepen was:

v.h.m.o. (Amsterdam, Den Haag)	53 meisjes 74 jongens	} 127 M
m.u.l.o. (Leiden)	61 meisjes 46 jongens	
n.o. (Gouda, Den Haag, Wassenaar)	85 meisjes 156 jongens	} 241 L
		475 proefpersonen

4.3. Statistische bewerking

Het aantal malen, dat éénzelfde item door één proefpersoon wordt gekozen, zullen wij in het vervolg aanduiden als de *itemscore*. Deze kan dus variëren tussen 0 en 4. Per item is nu nagegaan, in hoeverre significante verschillen in de frequentieverdeling van de itemscore tussen de proefgroepen optreden. Hierbij worden de items als onderling onafhankelijk behandeld.

Voorafgaand dient te worden onderzocht, in hoeverre deze laatste procedure geoorloofd is ten aanzien van aangrenzende items. Tussen deze dient immers te worden gekozen. Indien nu een item zeer sterke voorkeur krijgt, zodat de itemscore 4 bedraagt, impliceert dit, dat de score voor de horizontaal en verticaal aangrenzende items maximaal 2 kan zijn en die van de diagonaal aangrenzende items maximaal 3. Daardoor zullen bij deze items de frequentieverdelingen een geringere spreiding hebben en dus ook een geringer kans op significante verschillen geven. Met het oog op de constructie van een niveauschaal betekent dit, dat een bepaalde item, door zijn meer aantrekkelijke buurman als het ware uit de schaal wordt weggedrukt.

Dit verschijnsel behoeft, indien het van praktische betekenis zou blijken te zijn, nog niet als een argument tegen de gevolgde procedure te worden gebruikt. Het gaat er immers bij de schaalconstructie om, een voldoende aantal items te vinden, die op de gewenste wijze differentiëren. We behoeven voor de schaalconstructie niet alle potentieel geschikte items te gebruiken en in dit geval behoeft het verlies van een aantal „weggedrukte” items dus ook geen overwegend bezwaar te zijn, zolang een voldoende aantal andere overblijft. Desondanks is het interessant, zich een indruk te vormen van de betekenis van dit wegdruk-effect.

We hebben dit vraagstuk bestudeerd aan de hand van een niveauschaal van 20 items, in gelijke aantallen over de voor- en de achterzijde van het testformulier verdeeld. We beginnen met uit te gaan van één item, die tot de schaal behoort. Dit betekent, dat er van de 80 items op dezelfde bladzijde nog 9 tot dezelfde schaal behoren. De kans, dat een willekeurige item

daartoe behoort is dus $9/80$. Aan iedere item grenzen 8 andere, de kans, dat aan onze eerste item tenminste één andere grenst, die eveneens tot de schaal behoort, is derhalve $8 \times 9/80 = .90$. In totaal behoren tot de schaal 20 items. De kansverwachting dat aan een item uit de schaal tenminste één andere item grenst, die eveneens tot de schaal behoort, is echter niet $20 \times .9$, omdat dan in de kansberekening alle posities dubbel zouden worden geteld. De werkelijke kansverwachting is derhalve $20 \times .90/2 = 9$. In werkelijkheid bleek zich echter vijf maal het geval voor te doen, dat twee items uit deze schaal aan elkaar grensden, dus aanzienlijk minder dan verwacht. Indien dit verschil significant is, zou het er op kunnen wijzen, dat het wegdrukeffect inderdaad reële betekenis heeft.

Voor het onderzoek van de significantie van het verschil is door de heer NASS, hoofd van de afdeling Statistiek van het Nederlands Instituut voor Praeventieve Geneeskunde, een toets ontwikkeld, gebaseerd op de „Monte Carlo-methode” (¹⁰²). De kans, dat de gevonden afwijking van de verwachting door toevallige variaties wordt veroorzaakt, blijkt ca. 10% te bedragen, deze is dus te groot om de nulhypothese te verwerpen. We mogen hieruit wel concluderen, dat het wegdrukeffect, indien het al bestaat, praktisch van weinig betekenis zal zijn.

Voor de toetsing van de verschillen in frequentieverdeling op hun significantie, zijn de beide extreme niveaugroepen gebruikt. Aan deze werkwijze ligt vanzelfsprekend ten grondslag, dat eventueel voorkomende niveauverschillen zich het duidelijkst bij vergelijking van de n.o.-leerlingen met de v.h.m.o.-groep zullen openbaren. De karakteristiek van de m.u.l.o.-groep heeft eerst in een volgende fase van het onderzoek een rol gespeeld.

Bij de vergelijking hebben wij in eerste instantie de eventuele sexeverschillen buiten beschouwing gelaten. De totale v.h.m.o.-groep, voortaan als *M-groep* aan te duiden, is dus geplaatst tegenover de gehele groep leerlingen van het lager nijverheidsonderwijs (de *L-groep*).

Later zal blijken, dat er ten aanzien van de niveauvoorkeur wel enige sexeverschillen bestaan. Deze komen echter op ieder niveau voor en zijn telkens van dezelfde aard. Aangezien bovendien de percentages meisjes in de L- en M-groepen van dezelfde grootte-orde zijn (respectievelijk 35 en 42%), wordt het sexeverschil uit de intergroepvariatie grotendeels geëlimineerd. De door ons gevolgde procedure, die gemotiveerd was door een streven, de berekeningen niet onnodig omvangrijk te maken, kan dus achteraf gerechtvaardigd worden.

De frequentieverdelingen van de itemscores zijn als regel asymmetrisch. De gemiddelde itemscore bedraagt 1, zodat in de meeste gevallen een zeer geprononceerde linkszijdige scheefheid aanwezig moet zijn. Bovendien zijn de verdelingen, die wij willen vergelijken, in de meeste gevallen zelfs niet in eerste benadering isomorf. Voor de berekening van de significantie, tussen de verschillen in frequentieverdeling, moet daarom een distributievrije

toets worden gebruikt. We hebben hiervoor de toets van WILCOXON gekozen. Deze toets is elders uitvoerig beschreven (⁷¹, ¹¹²), zodat een gedetailleerde uiteenzetting hier achterwege kan blijven.

De toets van Wilcoxon is gebaseerd op de berekening van een toetsingsgrootheid U . Indien wij twee groepen personen vergelijken, waarvan de aantallen respectievelijk N_1 en N_2 bedragen ($N_1 + N_2 = N$), dan geldt voor U , dat het populatiegemiddelde en de variantie gelijk zijn aan $\mu = \frac{1}{2}N_1N_2$ $\sigma^2 = N_1N_2(N + 1)/12$. Deze grootheid U kan op zijn significantie worden onderzocht, doordat $T = (U - \mu) / \sigma$ bij voldoende grote waarden van N_1 en N_2 ($N_1 > 10$) steeds meer een standaard-normale verdeling gaat benaderen.

Indien in de beide frequentieverdelingen gelijke waarden voorkomen, hetgeen bij ons onderzoek als regel het geval is, geeft deze toetsing iets te geringe uitkomsten betreffende de significantie. Dit kan worden gecorrigeerd door een meer gecompliceerde toetsingsformule te gebruiken. In dat geval moet voor iedere item, die getoetst wordt, de berekening van μ en σ afzonderlijk geschieden, terwijl daarentegen de boven weergegeven formules voor alle items dezelfde waarden geven. Aangezien wij alleen geïnteresseerd zijn in duidelijk significante verschillen in de frequentieverdelingen, is een exacte toetsing echter niet noodzakelijk. We hebben daarom de bovengenoemde eenvoudige toets toegepast. De verschillen, die wij constateren, zijn dus in werkelijkheid nog iets signifikanter, dan hier aangegeven.

Bij ons onderzoek is N_1 (L-groep) = 241, N_2 (M-groep) = 127, dus $N = 368$. Hieruit volgt:

$$\mu = 15\ 303.5 \quad \sigma^2 = 941\ 165.25 \quad \sigma = 970.14$$

Voor alle 162 items werd de toetsingsgrootheid U berekend. Doordat deze grootheid lineair met T samenhangt, heeft ook U een normale verdeling, rondom een gemiddelde van 15 303.5, met een standaarddeviatie van 970.14. Gezien de grootte van de beide proefgroepen mag worden verwacht, dat de benadering van de normale verdeling zeer goed zal zijn. De wijze waarop U werd berekend, leidt er toe, dat items, die meer door de M-groep dan door de L-groep worden geprefereerd een hoge waarde van U zullen hebben. Omgekeerd wijst een lage waarde van U er op, dat de gemiddelde itemscore bij de L-groep hoger is dan bij de M-groep. Tabel 4.1 geeft voor alle 162 items de waarde van U . De eventuele significantie van het verschil is aangegeven met $*(P < .05)$, $** (P < .01)$ of $*** (P < .001)$. (Tabel 4.1, pag. 63).

4.4. Niveauverschillen in de bestaande schalen

Uit tabel 4.1 blijkt duidelijk, dat er niveauverschillen in de belangstelling bestaan en dat deze zelfs bij meer dan de helft van de items op significante wijze aan het licht treden. In totaal werd 84 maal de 5%-significantiedrempel overschreden, hiervan blijken 56 items significant op

TABEL 4.1. *U*-waarden der verschillen tussen de L- en de M-groep

1. 15171	42. 14879	82. 14284	123. 10722***
2. 13391*	43. 13177*	83. 13468	124. 19307***
3. 18885***	44. 10414***	84. 18156**	125. 10809***
4. 11358***	45. 14735	85. 12903*	126. 10616***
5. 13041*	46. 14929	86. 14648	127. 19238***
6. 17693*	47. 13372*	87. 19523***	128. 15980
7. 13135*	48. 14648	88. 12548**	129. 16212
8. 9665***	49. 18506***	89. 13997	130. 21381***
9. 19763***	50. 13792	90. 15304	131. 14858
10. 14670	51. 14980	91. 15729	132. 15048
11. 12010***	52. 20867***	92. 13945	133. 17649*
12. 12995*	53. 16003	93. 16841	134. 13776
13. 14033	54. 15911	94. 14405	135. 15766
14. 14178	55. 13912	95. 11406***	136. 14096
15. 12957*	56. 19955***	96. 13646	137. 19476***
16. 12968*	57. 12954*	97. 12591**	138. 11708***
17. 16930	58. 12824**	98. 16193	139. 12321**
18. 13416	59. 19481***	99. 13461	140. 18444***
19. 13584	60. 14533	100. 14508	141. 12900*
20. 15254	61. 13763	101. 14010	142. 14073
21. 20715***	62. 16708	102. 19029***	143. 20185***
22. 14180	63. 14530	103. 13211*	144. 11499***
23. 16219	64. 14376	104. 16449	145. 13319*
24. 18588***	65. 13595	105. 18122**	146. 12609**
25. 16217	66. 14545	106. 15420	147. 14899
26. 11159***	67. 12723**	107. 14136	148. 16435
27. 18076**	68. 13182*	108. 16305	149. 14046
28. 17860*	69. 13350*	109. 18448***	150. 11250***
29. 12023***	70. 12975*	110. 14878	151. 15256
30. 13533	71. 12130***	111. 13868	152. 19218***
31. 18149**	72. 13040*	112. 15318	153. 13095*
32. 11833***	73. 14097	113. 13313*	154. 15218
33. 12165**	74. 19766***	114. 12345**	155. 19462***
34. 18001**	75. 13330*	115. 16702	156. 15552
35. 17904**	76. 16092	116. 15655	157. 12589**
36. 11861***	77. 19154***	117. 14518	158. 13983
37. 13018*	78. 15854	118. 11824***	159. 14857
38. 18873***	79. 16936	119. 13950	160. 15877
39. 12685**	80. 19247***	120. 14356	161. 17627*
40. 16747	81. 17432*	121. 13290*	162. 14298
41. 13061*		122. 15261	

tenminste 1% en 40 van hen zelfs op tenminste .1%. Vooral deze grote concentratie van items met zeer significante verschillen in de frequentieverdeling is belangwekkend. Dit opent het perspectief, dat het inderdaad mogelijk zal zijn, een niveau-belangstellingsschaal te construeren. Opvallend is, dat een aanzienlijke meerderheid van de significante verschillen in de richting van de L-groep ligt. Het aantal items met zeer significante verschillen is in de beide richtingen nagenoeg gelijk. Het onderscheid is dus

een gevolg van het feit, dat in de richting van de L-groep een veel groter aantal items op de 5%-drempel significant zijn, dan in de richting van de M-groep. De verklaring van dit verschijnsel blijkt, wanneer we nagaan, tot welke bestaande belangstellingsschalen de items behoren, die significant verschillen opleveren. Tabel 4.2 geeft een overzicht van het aantal items per schaal, dat ten gunste van de L-groep, respectievelijk de M-groep differentieert.

TABEL 4.2. Voorkeursverschillen tussen L- en M-groep

Schaal	aantal items met voorkeur van	
	L-groep	M-groep
Technische handarbeid (TH)	11	0
Ambachtelijke vormgeving (AV)	8	1
Techniek en natuurwetenschap (TN)	0	14
Voedselbereiding (VB)	8	2
Agrarische arbeid (AA)	8	0
Handel (HA)	10	1
Administratieve arbeid (AD)	2	1
Literaire en geesteswetenschappelijke arbeid (LG)	0	15
Sociaal werk en opvoeding (SO)	3	0

De verschillen tussen de L- en de M-groep blijken duidelijk consistent. De voorkeur van de M-groep gaat zeer overwegend uit naar de items van de technisch-natuurwetenschappelijke en van de literair-geesteswetenschappelijke schalen, dus de beide typisch „intellectuele” interesserichtingen. Zeer duidelijk blijkt deze zelfde voorkeur ook bij nadere bestudering van de enkele items, die de specifieke preferentie van de M-groep hebben, maar die niet tot de beide genoemde schalen behoren. In de schaal AV heeft de M-groep meer voorkeur voor: 35 „Landkaarten tekenen”; in de schaal VB voor 38 „In een zuivelfabriek de melk bacteriologisch onderzoeken” en 152 „Scheikundige onderzoeken doen over levensmiddelen”; in de schaal AD: 81 „Loonbelasting berekenen”. De enige item waarvoor de voorkeur op deze wijze niet verklaard kan worden is in de schaal HA 124 „Artisten tegen ongevallen verzekeren”. Deze item krijgt een zeer significante voorkeur, zodat het uiterst onwaarschijnlijk is, dat we met een „toevals-significantie” te maken hebben. Nochtans is aan de hand van een andere steekproef nagegaan, of deze item opnieuw significant differentieerde, hetgeen inderdaad het geval bleek te zijn. De rationale van deze voorkeur is nog niet duidelijk. De concentratie van de preferenties der M-groep in twee sectoren, zal ook wel de verklaring zijn van het verschijnsel, dat de meeste significante items zo hoge *U*-waarden hebben bereikt. Bij de L-groep is de voorkeur uitgesmeerd over een veel groter aantal sectoren, hier zal zich daardoor eerder het verschijnsel van de items met slechts een betrekkelijk bescheiden significantie voordoen.

De voorkeur van de L-groep blijkt uit te gaan naar de vier sectoren van handarbeid en voorts naar de handel. Ook de agrarische arbeid trekt de belangstelling, zij het ook dat de meeste items van deze schaal geen zeer hoge mate van significantie bereiken. Vermoedelijk zou dit laatste bij inschakeling in het onderzoek van een groep leerlingen van het lager land- of tuinbouwonderwijs, anders zijn geweest.

De beide resterende sectoren: administratieve arbeid en sociaal werk en opvoeding bezitten weinig items, waarop een differentiatie naar niveau optreedt. Bij de eerstgenoemde zal dit wel samenhangen met het feit, dat kantoorwerk voor de meeste jongeren zeer weinig aantrekkelijkheid bezit, zodat de betreffende items in het algemeen weinig worden gekozen. Bij de tweede schaal zou men kunnen vermoeden, dat de kinderen als regel nog niet voldoende rijpheid bezitten om zich tot deze sector sterk aangetrokken te voelen.

Hoewel men achteraf gezien kan zeggen, dat het patroon van de voorkeursverschillen geheel ligt in de lijn van een verwachting, die men ook op grond van algemeen bekende feiten had kunnen opstellen, is het toch nuttig een dergelijk onderzoek uit te voeren. Indirect is het reeds een zekere validatie van de veronderstellingen waarop de test is opgebouwd. Bovendien kan men a priori niet altijd voorspellen, welke items uit de schalen significant zullen differentiëren. Wil men door combinatie van items tot de constructie van een schaal komen, die zo goed mogelijk naar niveau differentieert, dan is een kwantitatief onderzoek, zoals hier verricht, onontbeerlijk.

4.5. *Schaalconstructie*

Op basis van de gegevens, die in tabel 4.1 verzameld zijn, is overgegaan tot de constructie van een niveauschaal. Het zou denkbaar zijn, alle items, die significant differentiëren naar niveau, in een dergelijke schaal op te nemen. Een schaal bestaande uit ruim 80 items is echter praktisch bezwaarlijk te hanteren, terwijl men bovendien mag verwachten, dat de weinig significante items ook weinig tot de differentiërende werking van de schaal bijdragen. Daarom verdient het aanbeveling, uit de significant differentiërende items een keuze te doen.

Teneinde de differentiërende werking zo groot mogelijk te maken, zijn alleen items gekozen, waarvan de waarde van U lager was dan 12 500 of hoger dan 18 000, d.w.z. een waarde, die tenminste ongeveer driemaal de standaarddeviatie van het gemiddelde verschilt. Bij de normale verdeling bedraagt de éézijdige overschrijdingskans bij 3σ nog slechts .1%. (*Opmerking:* de significanties van tabel 4.1 zijn berekend als tweezijdige overschrijdingskansen.)

De vraag, hoe met de thans geselecteerde items een niveauschaal moet worden berekend, voert ons tot een preliminaire kwestie, die betrekking heeft op de dimensies van het niveaubegrip. In een praktische formulering

luit de vraag: moeten wij één niveauschaal construeren, of meerdere? Ten gunste van ieder van deze alternatieven kunnen argumenten worden aangevoerd.

De gedachte slechts één schaal te construeren is op het eerste gezicht evident. Het niveaubegrip, zoals wij dat tot nu toe hebben gehanteerd, is impliciet als unidimensioneel gedacht. Er is dan ook geen aanleiding, meer dan één niveauschaal te construeren. Vermoedelijk is ook in de Verenigde Staten, bij het construeren van de „occupational level”-schaal van de STRONG-test, hiervan uitgegaan. Deze werkwijze heeft bovendien het voordeel van de eenvoud. Wel moet men dan nog een oplossing vinden voor de praktische moeilijkheid, dat in de schaal items zouden worden opgenomen, die in verschillende richtingen differentiëren. Een uitweg kan worden gevonden, door een gedeelte der items een negatief teken te geven, de op deze items betrekking hebbende itemscore wordt dus bij de berekening van de totaalscore van de itemscores der overige items afgetrokken. Eventueel zou men ook voor deze items het complement van de itemscore kunnen gebruiken, d.w.z. het aantal malen, dat de betreffende items *niet* zijn gekozen en deze complementaire score bij de normale score op de overige items tellen. Laatstgenoemde procedure heeft het voordeel, dat geen negatieve uitkomsten mogelijk zijn.

Tegenover deze opvatting kan echter de overweging worden geplaatst, dat het niveaubegrip, zoals het hier wordt gehanteerd, gebaseerd is op een populair-psychologische opvatting, waarin een beroep wordt gedaan op een fysisch model. Het is echter lang niet zeker, dat dit model inderdaad zodanig geldig is, dat het ook bij onze schaalconstructie als uitgangspunt mag dienen. In het bijzonder sinds de studies van STOFFER c.s. ⁽¹¹⁹⁾, is wel bekend, dat de veronderstelde unidimensionaliteit van vele psychologische variabelen veel te wensen overlaat. Het is dus niet uitgesloten, dat het fenomeen, dat in de populaire denkwijze wordt opgevat als een één-dimensioneel niveauverschil, in wezen toch meer van kwalitatieve, dan van kwantitatieve aard is. Ook in de inleiding tot dit hoofdstuk hebben wij er op gewezen, dat het thans aan de orde zijnde verschijnsel zeker kwalitatieve aspecten heeft, die van de kwantitatieve moeilijk te scheiden zijn. Indien het niveauverschil tussen de L- en de M-groep in hoofdzaak zou moeten worden opgevat als een anders-gericht-zijn, hebben we te maken met twee verschillende attitudes. Deze zullen ongetwijfeld samenhangen met, tendele één-dimensioneel op te vatten, verschillen in begaafdheid, schoolvorderingen en milieu, maar zij zelf mogen toch niet worden behandeld als de twee extremen van één dimensie.

De vraag, welke van de beide zienswijzen de juiste is, is niet zonder experimentele toetsing te beantwoorden. In principe zou dit kunnen geschieden met behulp van een procedure, die door GUTTMAN ^(118, 47) hoofdstuk 6) is ontwikkeld. Deze werkwijze is echter zo gecompliceerd, dat een praktische toepassing zeer bezwaarlijk is. Wij hebben daarom een andere

weg gevolgd, waarbij het probleem indirect wordt benaderd.

Ter toelichting moet er eerst aan worden herinnerd, dat wij op één punt in het voorgaande de veronderstelling van unidimensionaliteit reeds hebben gebruikt: toen bij het opsporen van differentiërende items de m.u.l.o.-groep buiten beschouwing werd gelaten. Dit geschiedde op grond van de overweging, dat de leerlingen van deze groep (de *U-groep*) in het algemeen een midden-positie tussen de L- en de M-groep innemen. De m.u.l.o.-leerlingen vormen een zeer gemêleerd gezelschap, waarin men in grote aantallen kinderen aantreft, die eerder in het voortgezet lager onderwijs thuishoren en daar ook dikwijls na enige tijd belanden, terwijl er anderzijds ook een aanzienlijke groep is, die naar instelling eerder te vergelijken is met de M-groep. Op grond van deze overweging zijn wij impliciet uitgegaan van de verwachting, dat de gemiddelde score van de U-groep tussen die van de L- en de M-groep zal liggen. We kunnen nu nagaan, in hoeverre deze verwachting door de praktijk wordt bevestigd.

Van alle items, die voldoen aan de significantie-eis, welke aan het begin van deze paragraaf werd gesteld, is de gemiddelde itemscore van de L-, U- en M-groep bepaald. Dit betreft in totaal 46 items. Het bleek, dat in 37 gevallen de gemiddelde score van de U-groep inderdaad tussen die van L- en de M-groep lag. Het materiaal bevat dus inderdaad wel een aanwijzing, dat we met een unidimensioneel niveaubegrip te maken kunnen hebben. Naar onze mening is deze aanwijzing voldoende sterk, om alleen die items in de schaalconstructie te betrekken, waarvan de gemiddelde itemscore der U-groep tussen die van de L- en van de M-groep ligt.

Is de grondgedachte, dat we met een unidimensioneel niveaubegrip te maken hebben, juist geweest, dan moeten we verwachten, dat de som van de itemscores van de items, die door de L-groep worden geprefereerd, duidelijk negatief moet correleren met de som van de itemscores der items, die meer door de M-groep worden gekozen.

Deze veronderstelling is gemakkelijk te toetsen. We beginnen daartoe met de constructie van *twee* schalen, met respectievelijk de items, welke door de L-groep en door de M-groep worden verkozen. Voor de constructie van deze schalen staan op grond van het voorgaande 37 items ter beschikking. Besloten werd, items die slechts zeer zelden worden gekozen, niet in de schaal op te nemen, omdat deze tot de praktische differentiatie slechts een zeer geringe bijdrage leveren. Items, waarbij ook voor de groep, die de grootste affiniteit toonde, de gemiddelde itemscore beneden 1.00 bleef, zijn niet opgenomen. Voor de schaalconstructie bleven aldus 32 items over: 20 voor een schaal van items, die door de M-groep worden geprefereerd (de *M-schaal*) en 12 voor een *L-schaal*. Tabel 4.3 geeft een overzicht van de items, die in deze schalen zijn opgenomen.

Bij bestudering van de tabel blijkt, dat vrijwel alle items een grote mate van „face validity” hebben. Slechts twee items maken hierop een uitzondering: 124 „Artisten verzekeren tegen ongevallen” en 150 „De klanten van

TABEL 4.3. Items opgenomen in de L- en M-schalen

Item-nummer	gemiddelde itemscore		
	L-groep	U-groep	M-groep
<i>L-schaal</i>			
4. Autolampen monteren	1.19	.82	.64
8. Mooie etalages maken	2.53	2.19	1.63
29. Felicitatiekaarten tekenen	1.25	.98	.72
32. Gouden armbanden smeden	1.04	.64	.60
33. Aan een draaibank werken	1.37	.74	.73
44. De keuken van een groot hotel leiden	2.07	1.95	1.23
71. Feestgerechten voor een bruiloftspartij bereiden	1.58	1.45	1.15
95. Handelen in goud en sieraden	1.01	.71	.45
123. Dennenbossen in goede staat houden	1.31	1.08	.71
125. Een banketbakkerij leiden	1.33	.87	.59
138. Krukassen voor auto's draaien	1.27	.69	.68
150. De klanten van een bank bedienen	1.05	.90	.46
<i>M-schaal</i>			
9. Het weer van verschillende steden vergelijken	.80	.91	1.37
21. Geschiedenisles geven	.67	.95	1.51
24. Korte verhalen schrijven	1.27	1.77	1.79
27. Buitenlandse romans vertalen	1.09	1.31	1.58
31. In een electrotechnisch laboratorium werken	1.40	1.56	1.96
49. Amateurtoneelvoorstellingen regisseren	1.22	1.67	1.86
52. Aan een volksuniversiteit cursussen geven over geschiedenis	.53	.75	1.54
59. Landmetingen voor een ondergrondse spoorweg verrichten	.92	1.03	1.73
74. Verhalen schrijven voor geïllustreerde tijdschriften	1.13	1.31	1.71
80. Sagen en sprookjes verzamelen	.48	1.10	1.25
87. Plantencellen onder een microscoop onderzoeken	.86	1.20	1.57
102. Korte verhalen schrijven voor kranten	.86	1.27	1.42
109. De elektrische spanning van de bliksem meten	.99	1.00	1.53
124. Artisten verzekeren tegen ongevallen	.64	.83	1.28
130. Humoristische verzen maken	.54	.85	1.42
137. De nauwkeurigheid van thermometers testen	.52	.72	1.20
140. Nieuwe toepassingen voor aardgas zoeken	.97	1.07	1.50
143. Werken op een meteorologisch instituut	.76	.89	1.46
152. Scheikundige onderzoeken doen over levensmiddelen	.88	1.14	1.46
155. Biografieën van beroemde mannen verzamelen	.47	.72	1.17

een bank bedienen". Over het eerstgenoemde item hebben wij reeds eerder een opmerking gemaakt. Item nr. 150 neemt slechts schijnbaar een uitzonderingspositie in: deze item behoort tot de commerciële activiteiten,

waarvan in de vorige paragraaf reeds is gebleken, dat zij in het algemeen de voorkeur van de L-groep hebben.

Het eigenlijke statistische onderzoek van de beide schalen en ook de ijking zal pas in volgende hoofdstukken, in samenhang met de bewerking der overige schalen, aan de orde komen. Voorafgaand aan deze bewerking zijn echter over de L- en M-schaal reeds enkele studies uitgevoerd, die een voorlopige ijking beoogden. Met behulp van deze ijkingsgegevens was het mogelijk, reeds vroegtijdig enig inzicht in de validiteit van de schalen te krijgen.

Basismateriaal voor de voorlopige ijking vormden de resultaten van de proefpersonen, die in de eerste experimentele opzet waren betrokken geweest. Zij zijn, noch naar leeftijd, noch naar de frequentie van hun aantallen een goede afspiegeling van de bevolking van de zesde klas, zodat de uitkomsten van de ijking inderdaad als voorlopig moesten worden opgevat (voor de definitieve normen wordt verwezen naar hoofdstuk VI).

Voor de ijking is het nodig, de frequentieverdeling van de som der item-scores te bepalen. Het materiaal is daartoe gesplitst in zes subgroepen: de L-, U- en M-groep, ieder onderverdeeld naar sexe. Deze laatste onderverdeling scheen wenselijk op grond van de aard van de items in de beide schalen. Het is niet waarschijnlijk, dat deze steeds op gelijke wijze de voorkeur van jongens en meisjes zullen trekken.

Tabel 4.4 geeft een overzicht van enkele kenmerken van de frequentieverdelingen, gesplitst naar de zes subgroepen. Behalve de grootte van de subgroepen wordt de ligging van de kwartielen gegeven. Deze gegevens zijn ook voor het totale materiaal vermeld, hoewel de betekenis daarvan niet groot kan zijn, gezien het feit, dat dit materiaal niet representatief is.

TABEL 4.4. Frequentieverdeling der scores op de beide niveauschalen, naar opleidingsgroep en sexe (voorlopige ijking)

Groep	Meisjes				Jongens			
	N	Q ₁	Md	Q ₃	N	Q ₁	Md	Q ₃
<i>L-schaal</i>								
L-groep	85	14	17	21	156	14	17	21
U-groep	61	10	12	16	46	10	13	16
M-groep	53	7	10	13	74	7	10	14
Totaal	199	10	14	18	276	10	14	19
<i>M-schaal</i>								
L-groep	85	14	16	21	156	13	19	25
U-groep	61	16	20	27	46	22	26	34
M-groep	53	23	27	33	74	25	32	42
Totaal	199	16	20	27	276	16	23	31

Uit tabel 4.4 blijkt, dat de frequentieverdeling bij de L-schaal vrijwel geen sexeverschillen vertoont; ten naaste bij is bij alle drie opleidingsgroepen de ligging van de drie kwartielen voor de beide sexe dezelfde. De opleidingsgroepen daarentegen vertonen wel onderlinge verschillen. Hier valt op, dat de variatiebreedte van de verdelingen, althans wat betreft de afstand van het eerste tot het derde kwartiel, slechts weinig uiteenloopt: de verdelingen zijn bij benadering isomorf. De ligging van de grenslijnen schuift daarbij telkens één kwartiel op: Q_1 van de L-groep correspondeert ongeveer met de mediaan van de U-groep en met Q_3 van de M-groep. De mediaan van de M-groep beantwoordt ongeveer aan Q_1 van de U-groep, terwijl Q_3 van de U-groep vrijwel op één hoogte ligt met de mediaan van de L-groep. Door 25% van de M-groep en 25% van de L-groep te elimineren, zouden we op de twee extreme niveau's groepen krijgen, die elkaar vrijwel niet meer overlappen. We dienen er echter rekening mee te houden, dat deze conclusie is verkregen bij dezelfde steekproef, waarop ook de schalen zijn geconstrueerd. Bij kruisvalidatie moeten we een grotere mate van overlapping verwachten.

Teneinde te verifiëren hoe groot deze overlapping bij vergelijking van twee nieuwe steekproeven is, hebben we intermediair een normenschaal geconstrueerd. Daarbij is gebruik gemaakt van een transformatie op een rechthoekige verdeling, hetgeen bij interestests veelvuldig gebruikelijk is. Bij de rechthoekige verdeling is de kansverwachting voor iedere frequentie dezelfde, hetgeen meebrengt, dat ook de frequentie van de uiterste waarden hoog is. Bij een interestest zijn we juist in deze waarden geïnteresseerd. De standaarddeviatie van een rechthoekige verdeling is $1/\sqrt{12}$ maal de breedte van de rechthoek, zodat ongeveer 42.5% van de populatie een score heeft, die meer dan 1σ van het gemiddelde afwijkt. Bij de normale verdeling is dit bij slechts ca. 31.5% het geval.

Als regel wordt bij een interestest een percentielverdeling gebruikt. Aangezien echter de maximale score bij de grootste van onze beide schalen niet meer dan 74 punten bedroeg, hebben wij volstaan met een verdeling van 25 gelijke delen („vigintiquintiel”-verdeling). De normen zijn uiteraard op de totale groep bepaald, bij de L-schaal zijn ook de sexen tezamen genomen, terwijl bij de M-schaal afzonderlijke normen voor jongens en meisjes zijn berekend.

Uiteraard blijft het dan nog een open vraag, of dezelfde vigintiquintiel-score („V-score”) bij jongens en meisjes dezelfde betekenis heeft. Deze vraag zou alleen te beantwoorden zijn, indien in de eerste plaats reeds gegevens bekend waren over de predictieve validiteit van de schaal en bovendien bekend was of, onder overigens gelijke omstandigheden, de prognose ten aanzien van het studiesucces in de verschillende vormen van voortgezet onderwijs, voor jongens en meisjes gelijk is. Op geen van beide vragen is het antwoord met zekerheid bekend, terwijl de beantwoording van de tweede in ieder geval buiten het bestek van deze studie valt.

Aangezien de V-scores alleen voor intern gebruik bedoeld waren, zullen de normenschaal hier niet worden weergegeven. We vermelden alleen het resultaat op de L- en M-schaal bereikt door twee nieuwe proefgroepen, respectievelijk bestaande uit 100 leerlingen van het v.h.m.o. (jongens en meisjes, begin eerste leerjaar) en 100 leerlingen van het l.t.o. (jongens, eerste leerjaar).

De scores van deze leerlingen en die van de m.u.l.o.-leerlingen, die in de steekproef van de eerste experimentele opzet waren opgenomen, zijn vervolgens verwerkt in een cumulatieve frequentieverdeling. Langs grafische weg werden tussen de punten van deze frequentieverdelingen de best-passende curven bepaald. Bij analyse bleek, dat op de L-schaal de overlapping tussen de L- en M-groep ca. 25% bedroeg, d.w.z. ongeveer evenveel als bij het ijkingsmateriaal. De overlapping van de U- met de L-groep bedroeg ca. 40%, die van de U- met de M-groep ca. 30%. Op de M-schaal was de overlapping tussen de L- en de M-groep opnieuw 25%. De overlapping van de U- met de L-groep was nu ca. 30%, die van de U- met de M-groep ongeveer 45%. Door vergelijking met de gegevens van tabel 4.4 blijkt, dat op de M-schaal de afstand tussen de U- en de M-groep kleiner is geworden (vgl. de afbeeldingen 4.1 en 4.2).

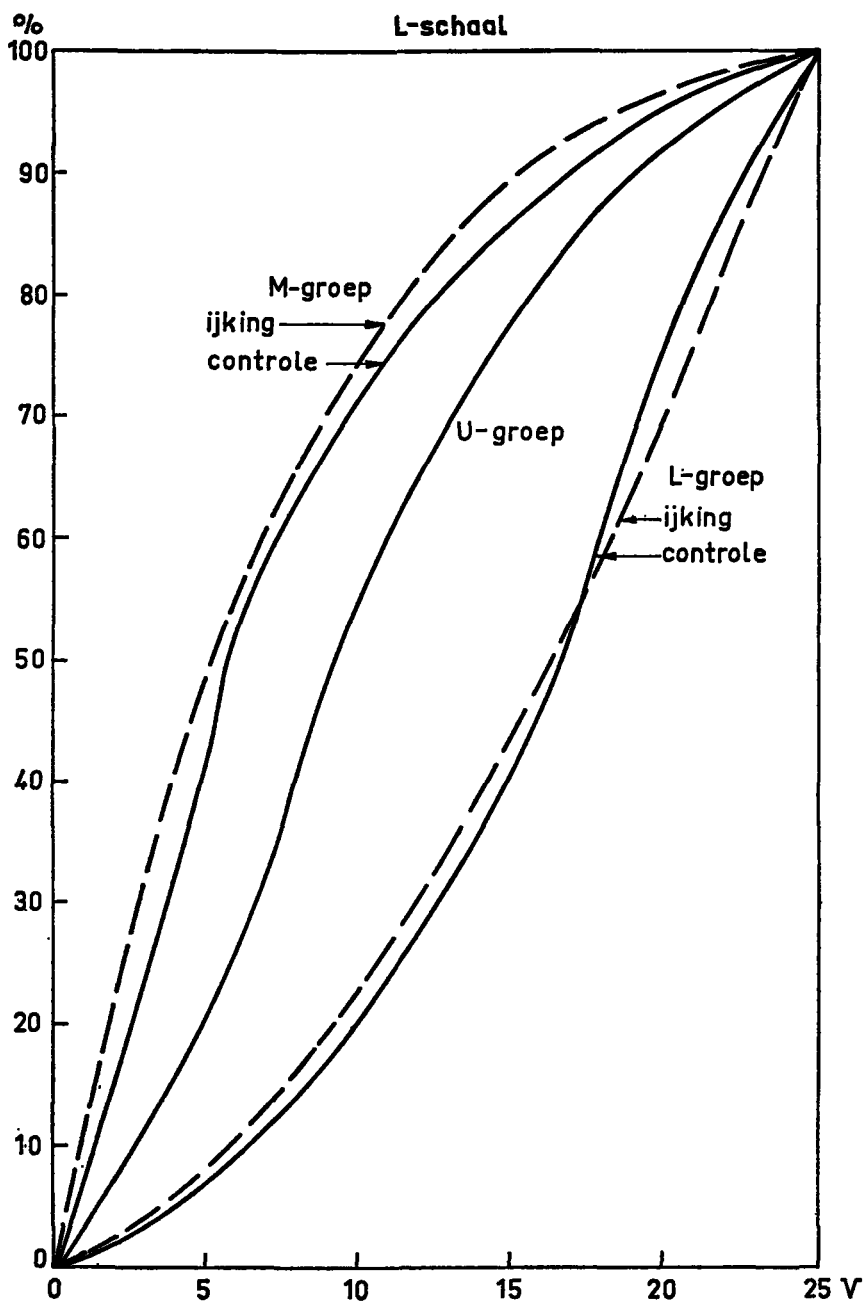
De bovenstaande resultaten wijzen er op, dat de beste differentiatie tussen het lager en het m.u.l.o.-niveau wordt verkregen met behulp van de M-schaal, terwijl de L-schaal de beste differentiatie tussen de m.u.l.o. en de v.h.m.o.-groep geeft.

Vervolgens is een eerste validatie van de L- en M-schaal uitgevoerd. Dit geschiedde in het kader van een onderzoek van 152 kinderen, afkomstig van een aantal zesde klassen van lagere scholen in de bollenstreek. Van deze kinderen waren 25 bestemd voor het v.h.m.o., 59 voor het m.u.l.o. en 68 voor het lager nijverheidsonderwijs, overig vakonderwijs, v.g.l.o. of voor rechtstreekse intrede in de praktijk. Van deze drie subgroepen werden op de beide niveauschaal de gemiddelde V-score en de standaarddeviatie bepaald. Tabel 4.5 geeft hiervan een overzicht.

TABEL 4.5. Eerste validatie van de beide niveauschaal

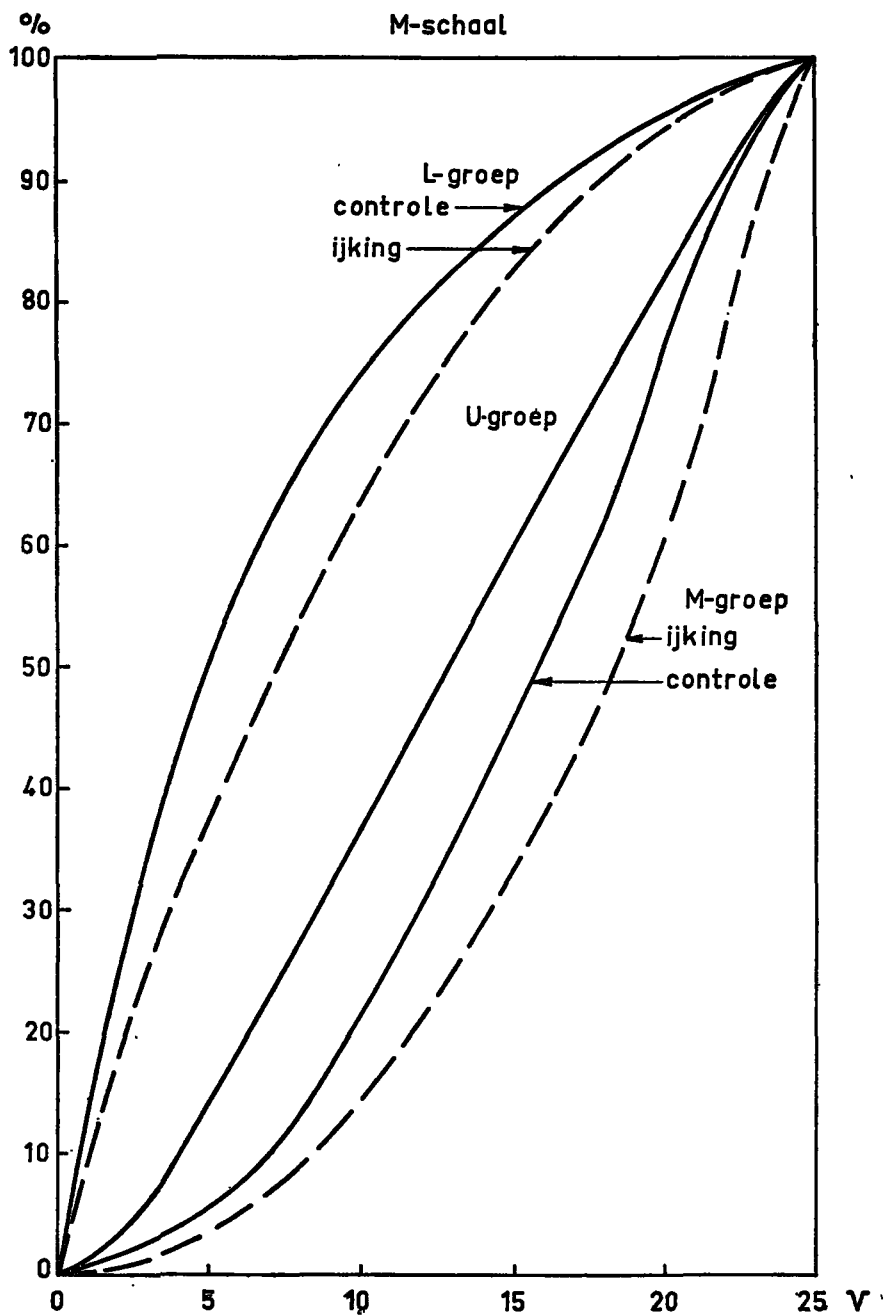
Groep	N	L-schaal		M-schaal	
		Gemidd.	s.d.	Gemidd.	s.d.
L-groep	68	17.3	4.9	7.6	5.1
U-groep	59	15.7	6.1	11.5	6.4
M-groep	25	12.5	7.0	12.8	6.5

De significantie van de verschillen kan worden bepaald door middel van de Student t-toets. Significant blijken op de L-schaal de verschillen tussen de L- en M-groep ($t_0 = 4.8$) en de U- en M-groep ($t_0 = 3.2$). Op de M-schaal zijn de verschillen tussen de L- en U-groep ($t_0 = 3.9$) en tussen de



Afbeelding 4.1

845



Afbeelding 4.2

846

L- en M-groep ($t_0 = 5.2$) significant. In al de genoemde gevallen wordt de 1%-drempel ruimschoots overschreden. De gevonden karakteristiek houdt dus in, dat de U-groep de tendentie heeft, op beide schalen betrekkelijk hoge gemiddelde scores te behalen. Daaruit volgt opnieuw, dat de differentiatie tussen de L- en U-groep het beste met behulp van de M-schaal en die tussen M- en U-groep het beste met behulp van de L-schaal kan geschieden. In ieder geval blijkt het mogelijk, tussen de drie niveau-groepen op bevredigende wijze een niveaudifferentiatie aan te brengen. Deze uitkomsten kunnen als een eerste validatie van de niveauschalen worden opgevat. De resultaten zijn zodanig, dat het verantwoord is, de beide schalen voor praktisch gebruik verder te bewerken.

Tenslotte moeten we nog de onderlinge samenhang van de L- en de M-schaal bezien. We zijn uitgegaan van de gedachte, dat, als er inderdaad sprake is van een één-dimensionaal niveauverschil, de L- en de M-schaal diametraal tegengesteld moeten differentiëren. In dat geval zou er dus, idealiter, een correlatie van -1.00 tussen de beide schalen moeten voorkomen. We hebben deze correlatie nagegaan aan een steekproef van 151 kinderen (67 jongens, 84 meisjes), allen leerlingen van zesde klassen van lagere scholen. Als correlatie werd gevonden: voor de jongens $r = -.584$; voor de meisjes $r = -.368$; voor de totale steekproef $r = -.412$.

Zelfs als men rekening houdt met de invloed van toevallige, storende factoren, die verhinderen, dat de optimale waarde bereikt wordt, blijven deze correlaties toch aan de lage kant. Dit kan dus een argument leveren ten gunste van de alternatieve opvatting, dat de zogenaamde niveauverschillen in werkelijkheid berusten op een anders-gericht-zijn. In ieder geval geven deze uitkomsten ons voorlopig aanleiding, de beide schalen afzonderlijk te handhaven.

4.6. *Tweede experimentele opzet*

Onafhankelijk van de hiervoor beschreven poging werd nog een tweede gedaan, het niveauprobleem te bewerken. De vraagstelling is ditmaal toegespitst op de selectie van leerlingen, die geschikt kunnen worden geacht voor het v.h.m.o. In hoeverre onderscheiden deze zich van de overige leerlingen van de lagere school? Het verschil kan liggen in een grotere toewending tot bepaalde activiteiten, maar ook in een afwending van andere activiteiten. In het eerste geval zullen de betreffende activiteiten, wanneer zij ter keuze worden gesteld, een relatief grote kans hebben, te worden gekozen. In het tweede geval wordt de betreffende activiteit, wanneer deze ter keuze wordt gesteld, vrijwel buiten beschouwing gelaten. Aangezien er echter telkens uit vier mogelijkheden moet worden gekozen, behoeft dit niet te betekenen, dat nu één andere item een sterke voorkeur krijgt. Het kan ook gebeuren, dat de voorkeur in dit geval vrijwel gelijkelijk over de resterende mogelijkheden wordt verdeeld.

Dit betekent strikt genomen, dat we beter doen, niet de items zoals

deze tot nu toe zijn gebruikt, als testeenheden te beschouwen, maar de kruispunten. Per kruispunt zijn vier keuzemogelijkheden gegeven, ieder van hen kan nu worden geanalyseerd, om na te gaan of er eventueel hetzij een significante toewending, hetzij een significante vermijding kan worden geconstateerd.

De verwerking van de zo verkregen bevindingen kan nog weer op verschillende wijzen geschieden. In de eerste plaats is het mogelijk, de diverse significante verschillen, die bij één kruispunt voorkomen, onderling te vergelijken op hun differentiërende werking. De meest effectieve kunnen dan in een schaal worden verenigd. De tweede mogelijkheid is, de verschillende significante verschillen een gewichtscijfer te geven, zodat ieder van hen in de waardering kan meespelen. Bij een schaal, die op deze wijze wordt geconstrueerd, kunnen we a priori een grotere differentiërende werking veronderstellen, dan bij de eerstbedoelde schaal. Hier staat echter tegenover, dat het gebruik van de schaal practisch bezwaarlijker is.

Aan het probleem van het toekennen van gewichtscoefficienten aan significante verschillen is in de literatuur vrij veel aandacht besteed. GULLIKSEN ⁽⁵⁸⁾, die aan dit probleem een uitvoerige beschouwing heeft gewijd, toont in de eerste plaats wel aan, dat er veel wegingsmogelijkheden zijn, zonder dat één daarvan uitgesproken de voorkeur verdient. Zowel de keuze van de methode, als die van de concrete gewichten, is sterk subjectief bepaald.

Bij belangstellingstests heeft het wegingsprobleem reeds relatief vroeg een rol gespeeld. Al in de twintiger jaren hebben o.a. COWDERY en STRONG bij hun interessentest met dergelijke gewichten gewerkt ⁽¹²⁴⁾, pag. 607 e.v.).

Hun procedure berust op de volgende berekening: We vergelijken twee groepen personen, de ene groep representatief voor de categorie waarvoor we een schaal willen construeren, de andere representatief voor de categorie, die we van de eerste willen onderscheiden (in het algemeen de doorsnee bevolkingsgroep). Deze groepen duiden we respectievelijk aan met I en II. We nemen nu de reacties op één concrete item. Er zijn verschillende reacties mogelijk (hoeveel doet niet ter zake), die we achtereenvolgens bestuderen. Neem hiervan één reactie. We vinden, dat van groep I, a% deze reactie heeft gegeven en b% (= 100—a) een andere reactie. Voor groep II zijn deze percentages respectievelijk c% en d%. De gewogen itemscore wordt nu:

$$\text{Itsc.} = 10 \frac{\varphi}{(1-\varphi^2) \sigma}$$

waarbij

$$\varphi = \frac{a-c}{\sqrt{(a+c)(b+d)}} \text{ en } \sigma = \frac{1}{2} \sqrt{(a+c)(b+d)}$$

De grootheid φ in deze formules kan gemakkelijk worden herkend als een punttetrachorische correlatie; σ is de standaarddeviatie van een dichotome proportionele verdeling. De factor $(1-\varphi^2)$ is evenredig met het

kwadraat van de standaard-schattingsfout. De factor 10 is bedoeld om decimalen te vermijden, zodat met gehele gewichtscoefficienten kan worden gewerkt. Uit de formules blijkt, dat de score een positief gewicht krijgt als $a > c$, terwijl in het omgekeerde geval een negatief gewicht wordt gegeven.

Bovenstaande formule is afkomstig van KELLEY, die echter later tot de conclusie is gekomen, dat de term in de noemer niet evenredig met de standaardfout van φ , maar met die van φ/σ behoorde te zijn. Hij heeft toen een andere, aanzienlijk ingewikkelder wegingsformule gegeven ⁽⁷⁴⁾, die door STRONG bij de laatste bewerking van zijn test is overgenomen.

GUILFORD ^(57a) heeft betoogd, dat het gebruik van de term $(1-\varphi^2)$ in de noemer van de eerstgenoemde formule onjuist is, omdat hij van mening is, dat de betrouwbaarheid van de regressiegewichten geen invloed op de bijdrage van de item tot de totale score behoort te hebben. Op grond hiervan komt hij tot een nieuwe, eenvoudige gewichtsformule, die tot op enkele constanten gelijk is aan φ/σ . ROCK en WESMAN hebben bij een vergelijking van de resultaten, die met behulp van de verschillende wegingsprocedures worden verkregen, gevonden dat zij onderling vrijwel gelijkwaardig zijn ⁽¹⁰⁹⁾.

KUDER, die bij de jongste vorm van de *Preference Record* eveneens van weging gebruik maakt, heeft in zijn „Research Handbook” ⁽⁸¹⁾ een gemakkelijk te volgen handleiding voor de berekening van al dan niet gewogen scores gegeven. Daarbij werd gebruik gemaakt van nomogrammen. De door hem gebruikte wegingsformule is zeer eenvoudig ⁽⁷⁹⁾:

$$Itsc = \frac{10 a}{a + c}$$

Op grond van deze formule, die ook door ons werd gebruikt, kan de gewichtsfactor variëren tussen 0 ($a = 0\%$) en 10 ($c = 0\%$). Indien $a = c$ is de gewichtsfactor = 5. De gewichten worden op gehele getallen afgerond.

Voor de berekening van een „gewogen” MO-schaal werd een nieuwe steekproef gebruikt. Gezien de ervaringen bij het voorgaande onderzoek, scheen het gewenst, de gevallen van jongens en meisjes bij voorbaat afzonderlijk te behandelen.

Er werden daarom twee proefgroepen geconstrueerd, ieder bestaande uit 75 jongens en 75 meisjes. De MO-groep bestond uit kinderen, die in de eerste maanden van de eerste klas van een v.h.m.o.-school waren onderzocht en waarvan bovendien bekend was, dat zij aan het einde van het eerste jaar zonder taak of herexamen waren bevorderd. Deze groep bestond uit andere kinderen dan degenen, die in de eerste experimentele opzet betrokken waren geweest. De tweede groep was bedoeld als representant van de lagere-schoolbevolking in het algemeen. Zij bestond uit kin-

deren, die in de laatste maanden van de zesde klasse waren onderzocht. Deze groep vormde een onderdeel van de landelijke steekproef, die ten behoeve van de ijking van de B I T was gebruikt (vergelijk hoofdstuk VI). In verband echter met het feit, dat de laatste jaren ruim 10% van de schoolbevolking van het lager onderwijs naar het v.h.m.o. gaat, zou er bij voorbaat een aanzienlijke essentiële overlapping tussen de beide proefgroepen moeten worden verwacht. Besloten werd, deze te elimineren, door geen kinderen in de tweede proefgroep op te nemen waarvan bekend was, dat zij óf een school voor v.h.m.o. óf een „brugklasse” zouden gaan bezoeken. Op deze wijze wordt een scherpere differentiatie mogelijk.

Volgens de procedure van KUDER werd eerst nagegaan, welke reacties op ieder kruispunt een significant verschil tussen de beide proefgroepen opleverden. De berekening geschiedde voor jongens en meisjes afzonderlijk. Met behulp van een door KUDER gegeven nomogram werd vervolgens de gewichtsfactor vastgesteld. Deze kan zoals gezegd variëren tussen 0 en 10. De waarden 0, 1, 9 en 10 bleken echter slechts zeer zelden voor te komen, terwijl verschillen met het gewichtscijfer 4 of 6 dikwijls nauwelijks significant waren. Daarom werd de scoring vereenvoudigd: de waarden 0 t/m 2 werden als + 2 gerekend, de waarde 3 als +1, de waarden 4 en 6 als 0, de waarde 7 als -1 en de waarden 8 t/m 10 als -2.

Een dergelijke vereenvoudiging ligt geheel in de lijn van de ontwikkeling, die b.v. ook bij de STRONG-test heeft plaatsgevonden. Aanvankelijk heeft men met zeer grote gewichten gewerkt, bij COWDERY varieerden deze tussen +30 en -30, in de eerste editie van de STRONG-test waren de variatiegrenzen +15 en -15. Later zijn zij aanzienlijk gereduceerd. Bij de thans in gebruik zijnde uitgave van de *Vocational Interest Blank* worden gewichten tussen +4 en -4 toegepast. STRONG⁽¹²⁴⁾ meent, dat met laatstgenoemde waarden een optimaal resultaat wordt verkregen, maar erkent, dat een verdergaande reductie van de gewichten, zelfs tot de eenheid toe, niet veel slechtere uitkomsten geeft.

De praktische betekenis van deze vereenvoudiging blijkt onmiddellijk, wanneer wij de scoringsprocedure beschrijven. Voor iedere gewogen waarde moet een afzonderlijke nakijkschablone worden vervaardigd. Deze bestaat uit twee bladen: voor de beide zijden van het testformulier. Zij zijn voor de beide sexen verschillend. Na de vereenvoudiging zijn er nog vier verschillende gewogen waarden (de waarde 0 hoeft uiteraard niet te worden gescord) zodat voor de scoring toch nog 8 bladen moeten worden gehanteerd. Indien men echter met gewogen scores van 0 t/m 10 werkt, zouden 20 bladen nodig kunnen zijn!

Onafhankelijk van het bovenstaande is ook nog een berekening uitgevoerd, die tot doel had, per item een gewogen score te bepalen. In eerste instantie gelijkt deze procedure sterk op die, welke in de eerste experimentele opzet werd gevolgd; alleen is nu de significantie van de verschillen berekend volgens de methode van KUDER. Eveneens volgens KUDER

zijn vervolgens de gewichten berekend, die daarna op de hiervoor beschreven wijze werden vereenvoudigd. Voor meisjes en jongens afzonderlijk zijn zodoende schalen verkregen, gebaseerd op gewogen item-scores.

De tabellen 4.6 en 4.7 geven een overzicht van de aantallen kruispuntreacties, c.q. items, die in de schalen zijn opgenomen. De gegevens zijn gesplitst naar sexe en naar de voor- en achterzijde van het testformulier.

TABEL 4.6. Aantal kruispuntreacties in gewogen MO-schaal

Jongens		gewogen score	Meisjes	
Voorkant	Achterkant		Voorkant	Achterkant
12	7	+ 2	13	8
35	29	+ 1	10	12
9	10	0	9	15
—	—	—	—	—
56	46	totaal positief	32	35
26	13	— 2	10	12
20	18	— 1	6	8
5	2	0	5	12
—	—	—	—	—
51	33	totaal negatief	21	32

Opmerking: De 0-scores spelen alleen een rol, indien men een schaal zou willen gebruiken, waarin de kruispuntreacties ongewogen worden verwerkt. In dat geval worden de voorkeursreacties uitsluitend naar teken gesommeerd.

TABEL 4.7. Aantal items in gewogen MO-schaal

Jongens		gewogen score	Meisjes	
Voorkant	Achterkant		Voorkant	Achterkant
1	0	+ 2	3	2
8	6	+ 1	0	4
2	4	0	0	0
—	—	—	—	—
11	10	totaal positief	3	6
3	0	— 2	1	1
8	1	— 1	2	6
1	0	0	0	0
—	—	—	—	—
12	1	totaal negatief	3	7

De sexeverschillen blijken dus inderdaad aanzienlijk te zijn, met name blijkt er ook een groot verschil in voorkeur ten opzichte van respectievelijk de items aan de voor- en achterzijde te bestaan. Ten aanzien van de items is nog eens in detail nagegaan, welke de voorkeursverschillen tussen jongens en meisjes zijn. De betreffende gegevens zijn in de tabellen 4.8 en 4.9 verwerkt.

TABEL 4.8. Items, die zijn opgenomen in een gewogen MO-schaal voor meisjes

	Gewicht
Karpervijvers aanleggen	+ 2
Onderzoek doen naar nieuwe steenkoolgebieden	+ 2
Landmetingen voor een ondergrondse spoorweg verrichten	+ 2
Forellen vangen	+ 2
Orchideeën kweken	+ 2
Rashonden fokken	+ 1
Plantencellen onder een microscoop onderzoeken	+ 1
Humoristische verzen maken	+ 1
Een bijenteelt beginnen	+ 1
Verkoopbrieven schrijven	— 1
Koffie inkopen voor een koffiebrandery	— 1
Goede verkoopmanieren leren aan verkopers	— 1
Klanten raad geven bij de aankoop van meubelen	— 1
Artisten tegen ongevallen verzekeren	— 1
Een banketbakkerij leiden	— 1
In een proefkeuken diëtrecepten proberen	— 1
De klanten van een bank bedienen	— 1
Motorfietzers tegen ongevallen verzekeren	— 2
Handelen in goud en sieraden	— 2

Opvallend is hier vooral, dat onder de items met positieve gewichtsfactor vrijwel geen representanten van de literair-geesteswetenschappelijke schaal voorkomen. Het sterkst vertegenwoordigd zijn de items uit de agrarische sector, waarbij opvalt, dat de voorkeur zich concentreert op de „luxe” activiteiten. Daarnaast zijn enkele items uit de natuurwetenschappelijke sector in trek. De items met negatieve gewichtsfactoren komen vrijwel uitsluitend uit de commerciële sector. Zelfs nr. 124, „artisten tegen ongevallen verzekeren”, die bij het eerste onderzoek en ook in een contrôle-steekproef de preferentie van de M-groep had, wordt hier afgewezen. Twee afgewezen items komen uit de voedingssector.

De voorkeur van de jongens verschilt in een aantal opzichten sterk van die van de meisjes. Van de 15 positief gekozen items behoren 13 tot de technisch-natuurwetenschappelijke sector, terwijl een 14e, die officieel tot de voedingssector wordt gerekend, een uitgesproken natuurwetenschappelijke inslag heeft. Bovendien blijken van de 6 items, die een gewichtsfactor 0 hebben, er 5 eveneens tot de natuurwetenschappelijke sector te behoren, zodat in dit geval *alle* items van de schaal TN ten gúnste van de MO-groep blijken te differentiëren. De zesde item met gewichtsfactor 0: „landkaarten tekenen”, past alweer naar aard goed bij de natuurwetenschappelijke interesses. In negatieve zin valt op, dat de meeste items, die worden afgewezen, tot de agrarische sector behoren. Slechts in één geval betreft dit echter een item, dat door de meisjes positief wordt verkozen, de overige items hebben betrekking op „gewone” agrarische bezigheden. Verder wor-

Tabel 4.9. Items, die zijn opgenomen in een gewogen MO-schaal voor jongens

	Gewicht
Plastics onderzoeken op zuurbestendigheid	+ 2
De sterkte van ijzeren draagbalken berekenen	+ 1
Het weer van verschillende steden vergelijken	+ 1
Het traject voor een nieuwe spoorwegtunnel bepalen	+ 1
In een electrotechnisch laboratorium werken	+ 1
De haarden van aardbevingen bepalen	+ 1
Onderzoek doen naar nieuwe steenkoolgebieden	+ 1
Landmetingen voor een ondergrondse spoorweg verrichten	+ 1
Verhalen schrijven voor geïllustreerde tijdschriften	+ 1
Erfelijkheidsproeven op planten doen	+ 1
Plantencellen onder een microscoop onderzoeken	+ 1
Het prestatievermogen van motoren verbeteren	+ 1
De nauwkeurigheid van thermometers testen	+ 1
Nieuwe toepassingen voor aardgas zoeken	+ 1
Scheikundige onderzoekingen doen over levensmiddelen	+ 1
Eieren uitbroeden in een broedmachine	— 1
Rozen stekken	— 1
Beukenbossen uitdunnen	— 1
Gouden armbanden smeden	— 1
Motorfietsers tegen ongevallen verzekeren	— 1
Bossen aanplanten	— 1
Kasbloemen kweken	— 1
Karpervijvers aanleggen	— 1
Zwarte tulpen kweken	— 1
Auto's verkopen	— 2
Gereedschappen smeden	— 2
Tabakspijpen uit hout snijden	— 2

den afgewezen twee commerciële activiteiten, twee items, die op smeedwerk betrekking hebben en nog een, enigszins buitenissig, item uit de sector der ambachtelijke vormgeving.

De voorkeursreacties van de jongens passen over het geheel beter bij de items, die in de eerder beschreven M- en L-schalen zijn opgenomen, dan die van de meisjes. Opvallend is, dat hier de belangstelling voor de exacte onderwerpen nog veel meer geprononceerd is, dan in het geval van het eerste onderzoek. Ter verklaring willen we verwijzen naar het feit, dat de L-groep in het eerste geval, wat de jongens betreft, voornamelijk was samengesteld uit leerlingen, die een technische opleiding hadden gekozen, terwijl in dit tweede geval ook kinderen met andere beroepsperspectieven, b.v. in de agrarische sector, in de steekproef waren opgenomen. Dit verklaart tevens, waarom in negatieve zin de agrarische items uit de L-schaal zich beter hebben weten te handhaven dan de overige. Niet geheel duidelijk is, waarom zowel bij jongens als meisjes de literair-geesteswetenschappelijke items zo weinig differentiërende waarde blijken te bezitten, terwijl zij toch in de eerste M-schaal in grote getale vertegenwoor-

digd waren. De voorkeur van de v.h.m.o.-meisjes voor „luxe” items uit de agrarische sector, zal waarschijnlijk moeten worden geïnterpreteerd als een milieu-gebonden voorkeur, die ook in de liefhebberijensector veelvuldig tot uiting komt.

4.7. Intelligentie en interesse

We kunnen het niveaubegrip ook enigszins anders opvatten dan tot dusver is geschied en het betrekken op verschillen in verstandelijke begaafdheid. Deze nieuwe interpretatie staat uiteraard niet geheel los van de eerste, traditioneel is de verstandelijke begaafdheid één der voornaamste criteria, waarop de differentiatie voor het voortgezet onderwijs wordt gebaseerd. Onderzoek naar de relatie tussen intelligentie en interesse is reeds betrekkelijk spoedig na de constructie van de eerste belangstellingstests geschied. De meeste onderzoeken hebben tot de slotsom gevoerd, dat er slechts een geringe correlatie tussen de beide genoemde grootheden bestaat ^(20, 49).

Gezien deze over het geheel negatieve bevindingen hebben wij geen uitgebreid onderzoek over dit onderwerp gedaan, maar volstaan met gebruik te maken van een toevallig geboden mogelijkheid tot verificatie.

In het kader van een ander onderzoek zijn een 150-tal kinderen, leerlingen van zesde klassen, onderzocht, zowel met een intelligentietest (Groninger Afsluitingsonderzoek Lager Onderwijs) als met de B I T. Uit dit materiaal zijn aselekt de gegevens betrekking hebbende op 50 jongens en 50 meisjes genomen, voor een berekening van de correlatie tussen het I.Q. en de ruwe waarde op de negen schalen van IRLE. Bovendien is op het totale materiaal de correlatie tussen het I.Q. en de beide door ons geconstrueerde schalen L en M berekend. Tabel 4.10 geeft een overzicht van de uitkomsten van de berekening.

TABEL 4.10. Correlaties tussen intelligentieniveau (I.Q.) en belangstelling

schaal	meisjes	jongens	totaal
Technische handarbeid	—,364	—,257	.045
Ambachtelijke vormgeving	.072	—,172	—,144
Techniek en natuurwetenschap	.003	.269	.258
Voedselbereiding	—,191	—,166	—,258
Agrarische arbeid	.225	.018	.164
Handel	—,471	.180	—,088
Administratieve arbeid	—,218	.231	.108
Literatuur en geesteswetenschappen	.275	.151	.139
Sociaal werk en opvoeding	.075	—,162	—,153
M	.192	.327	.255
L	.314	—,206	.071

De meeste waarden uit deze tabel blijken niet significant van nul te verschillen, waarmee de conclusie van de Amerikaanse onderzoekers ook

voor de B I T wordt bevestigd. Enkele opvallende punten kunnen nog kort worden gereleveerd.

In de eerste plaats hebben de totaalwaarden in verschillende gevallen waarschijnlijk weinig betekenis. Dit blijkt wel uit het feit, dat zij soms zeer veel verschillen van het gemiddelde van de afzonderlijke waarden voor de sexen (vergelijk b.v. de technische handarbeid). Dit is een gevolg van verschillen in de frequentieverdelingen bij jongens en meisjes.

De zeer sterk negatieve correlatie van commercieel werk met intelligentie bij meisjes komt overeen met de uitkomsten van het onderzoek, dat in de voorgaande paragraaf werd beschreven, hetzelfde geldt voor de positieve correlatie met agrarische arbeid. Opvallend is dan wel, dat de sterk negatieve correlatie met technische handarbeid, niet in de gewogen MO-schaal tot uitdrukking is gekomen, zeer waarschijnlijk ten gevolge van het feit, dat de betreffende items bij meisjes in het algemeen weinig worden gekozen. Bij meisjes blijkt de L-schaal niet op de verwachte wijze te differentiëren, in tegenstelling tot de jongens. Het feit, dat bij jongens de agrarische schaal geen duidelijk negatieve correlatie met het I.Q. heeft, moet waarschijnlijk op externe factoren worden teruggevoerd: het hier beschreven onderzoek geschiedde in een omgeving, waar agrarische arbeid een grote rol speelt, terwijl de steekproef uit de vorige paragraaf een nationaal en dus meer urbaan karakter had. Overigens liggen de van nul verschillende waarden bij jongens geheel in de verwachte richting,

HOOFDSTUK V

DE PSYCHOMETRISCHE KWALITEITEN VAN DE B I T-SCHALEN

5.1. *Vraagstellingen en experimentele opzet*

De tegenwoordige testgebruiker stelt aan zijn instrument gewoonlijk een aantal eisen, die betrekking hebben op de psychometrische kwaliteiten. Er moeten bewijzen van deugdelijkheid van de test, voor de toepassing in bepaalde situaties of voor bepaalde doeleinden worden gedemonstreerd. Deze twee aspecten, de instrumentele deugdelijkheid en de gebruikswaarde, staan niet geheel los van elkaar. Maar de problematiek, die hiermee samenhangt vertoont toch zo veel verschilpunten, dat een gescheiden behandeling gewenst is.

Bij het onderzoek van de interne deugdelijkheid van de test, hebben we te maken met een beperkt aantal welomschreven kwaliteiten, waarvan de experimentele toetsing niet al te gecompliceerd is. Er doen zich hier wel vraagpunten voor, waarvan de oplossing niet steeds ondubbelzinnig mogelijk is, maar in hoofdzaak kan de toetsing toch geschieden volgens een schema, dat algemeen aanvaard wordt. Daarentegen is de toetsing van de externe waarde van de test veel meer problematisch. De hiermee samenhangende vraagstukken, die we gewoonlijk als het validatieprobleem aanduiden, zijn nog maar ten dele opgelost. Daarom zal in dit hoofdstuk alleen het onderzoek naar de interne deugdelijkheid van de test aan de orde worden gesteld. De beschouwingen over het validatieprobleem volgen eerst, wanneer ook de ijking van de test is besproken.

Als eerste aspect van de deugdelijkheid beschouwen we de homogeniteit der schalen. Willen we de interesse ten aanzien van een bepaalde belangstellingssector beoordelen, dan zal dit moeten geschieden met behulp van een instrument, dat innerlijk consistent is. Dat wil zeggen, dat we in de betreffende schaal items moeten hebben opgenomen, die stuk voor stuk op dezelfde wijze differentiëren, als de schaal dit doet. Degenen, die op de schaal een hoge score behalen, behoren gemiddeld ieder der items vaker te kiezen, dan degenen, die op de schaal een lage score behalen. Vanzelfsprekend is dit een statistische uitspraak, individueel zijn afwijkingen van de verwachting heel wel mogelijk. We noemen dit kenmerk de homogeniteit der items met de schaal. Een homogeniteitsanalyse is dus de eerste opgave van dit hoofdstuk. In het kader van dit homogeniteitsonderzoek zijn ten aanzien van één der schalen bezwaren aan het licht gekomen,

die ons aanleiding hebben gegeven, de betreffende schaal te herzien. Het verslag van deze revisie zal in aansluiting op de paragraaf over het homogeniteitsonderzoek worden gegeven.

Vervolgens zullen de intercorrelaties van de schalen moeten worden onderzocht, teneinde na te gaan, of we inderdaad met interessevormen te maken hebben, die in voldoende mate onderling onafhankelijk zijn.

Een derde belangrijk instrumenteel kenmerk is de betrouwbaarheid van de test. De hierop betrekking hebbende gegevens zijn in een slotparagraaf opgenomen.

Ten behoeve van de verschillende in dit verband verrichtte onderzoeken zijn diverse proefgroepen gebruikt. De verschillen in samenstelling hangen tendele samen met de technische eisen, die aan de verschillende berekeningen worden gesteld. Bij een toetsing, waaraan zeer veel rekenwerk is verbonden, zal men dikwijls uit praktische overwegingen met een kleiner materiaal volstaan, dan in gevallen, waarin de omvang van de steekproef ons vrijwel geen technische beperkingen oplegt. Bovendien zijn de verschillende hier gerapporteerde onderzoeken niet alle in hetzelfde stadium van het onderzoek geschied, ook de aard van het beschikbare materiaal heeft de samenstelling van de steekproeven beïnvloed. In iedere paragraaf zullen over de concrete samenstelling van het materiaal, waarop de berekeningen zijn gebaseerd, mededelingen worden gedaan.

5.2. *Homogeniteitsanalyse*

Het begrip *homogeniteit* doet nog maar betrekkelijk korte tijd opgeld in de test-theorie. Aanvankelijk heeft men de homogeniteit wel eens als een bijzondere vorm van betrouwbaarheid beschouwd en er als zodanig bezwaren tegen gemaakt ⁽⁹¹⁾, hoofdstuk 15). Deze bezwaren, die op een misvatting berusten, zijn daardoor niet relevant. Hoogstens kan men zeggen, dat een test, waarvan alle items een hoge mate van homogeniteit met de test hebben, waarschijnlijk ook volgens alle criteria een hoge graad van betrouwbaarheid zal hebben. Het omgekeerde is echter niet zonder meer waar.

Practische betekenis heeft het homogeniteitsbegrip vooral door het werk van LOEVINGER ⁽⁹²⁾ gekregen, die daarbij kon voortbouwen op een oudere studie van LONG ⁽⁹³⁾. Van de verschillende vormen van homogeniteit, die door LOEVINGER worden onderscheiden, is er slechts één in verband met onze studie van aanzienlijke betekenis. Deze betreft de homogeniteit van een item met een test- of schaalvariabele. PEAK ⁽¹⁰⁵⁾, die aan dit onderwerp een beschouwing heeft gewijd, is van oordeel, dat er een wezenlijke verwantschap bestaat tussen de methode van LOEVINGER en de onderzoekstechniek van GUTTMAN ⁽¹¹⁸⁾ over de dimensionaliteit van schalen. Naar haar mening verdient de werkwijze van LOEVINGER echter de voorkeur, omdat deze minder aanvechtbare punten heeft. Wij zouden hieraan willen

toevoegen, dat zij ook het niet te onderschatten voordeel van praktische hanteerbaarheid bezit.

Bij de bepaling van de homogeniteit wordt uitgegaan van de vergelijking van personen, die op de test of schaal als geheel (of beter, op de test minus de te onderzoeken item) verschillende scores behalen. Bij analyse van de itemscore kunnen we dan drie gevallen onderscheiden: a) homogene onderscheiding: degene met de hoogste testscore heeft ook de hoogste itemscore; b) heterogene onderscheiding: degene met de hoogste testscore heeft de laagste itemscore; c) geen onderscheiding: beide hebben dezelfde itemscore. We noemen nu een item te meer homogeen met de test- of schaalvariabele, naarmate de kans groter is, dat degene met de hoogste testscore ook de hoogste itemscore heeft, dus naarmate het percentage homogene onderscheidingen groter is.

Teneinde de homogeniteit in werkelijkheid te bepalen, zullen we alle combinaties van twee personen uit het materiaal ten aanzien van de itemscore moeten vergelijken; met dien verstande, dat men de combinaties, waarin beide personen dezelfde totaalscore behalen, buiten beschouwing pleegt te laten. Bij de bepaling van de totaalscore, wordt de onderzochte item niet meegeteld, teneinde een „spurious correlation” te vermijden.

LOEVINGER heeft een homogeniteitsindex geconstrueerd, die geheel gebaseerd is op de verhouding van de homogene en de heterogene onderscheidingen. De gevallen van „geen onderscheiding” worden buiten beschouwing gelaten. Aangezien we hier in feite met een bijzonder geval van correlatie te maken hebben, is een index gekozen, waarvan de waarde kan variëren tussen -1 en $+1$.

Noemen de H_o het aantal homogene onderscheidingen en H_e het aantal heterogene, dan luidt de formule voor de homogeniteitsindex:

$$H_{it} = 1 - \frac{2 H_e}{H_o + H_e}$$

Hieruit blijkt onmiddellijk, dat $H_{it} = 1$ als $H_e = 0$, $H_{it} = -1$ als $H_o = 0$, terwijl $H_{it} = 0$ als $H_o = H_e$. In de literatuur wordt wel eens aangevoerd, dat de betekenis van de tussenliggende waarden onduidelijk is. We menen echter, dat een interpretatie mogelijk is, wanneer men bedenkt, dat de homogeniteitsindex een sterke overeenkomst vertoont met de rangcorrelatie-index *tau* van KENDALL (⁷⁶); met dien verstande, dat in de homogeniteitsindex de gevallen van geen onderscheiding buiten beschouwing worden gelaten, hetgeen bij de index van KENDALL niet het geval is. Volgens DE JONGE (⁷¹) behoeft echter het weglaten van de „ties” niet als een ernstig bezwaar te worden geteld, zolang hun aantal betrekkelijk bescheiden is — hetgeen bij homogeniteitsonderzoek als regel van toepassing is. Men kan dan dus de homogeniteitsindex behandelen als een representant van KENDALL's *tau* en dienovereenkomstig op zijn significantie toetsen.

Zolang we te maken hebben met items, waarop slechts twee reacties mogelijk zijn, b.v. ja/nee, goed/fout of dergelijke, heeft de homogeniteitsindex geen voordelen boven andere correlatie-indices. BOON VAN OSTADE ⁽²¹⁾ heeft in een onderzoek, dat betrekking had op 50 testitems, gevonden, dat er een zeer hoge correlatie bestond tussen de waarde van de homogeniteitsindex en die van de biseriale correlatie. De product-momentcorrelatie tussen de beide reeksen waarden bedroeg .98. In dit geval zou men waarschijnlijk de biseriale correlatie prefereren, die voor het geval van twee alternatieven gewoonlijk sneller te berekenen is. De situatie verandert echter, zodra men te maken krijgt met items, waarop meerdere reacties mogelijk zijn. Bij intelligentietests kan zich dit geval b.v. voordoen, indien voor iedere item een gedifferentieerde puntenwaardering met 3 of 4 schaalpunten wordt gebruikt (vergelijk ¹¹⁷). Ook bij onze test, waarbij iedere item 0—4 maal kan worden gekozen, treedt dit geval op. Weliswaar bestaan er ook formules voor sommige vormen van multiseriale correlatie, tot quintiseriale correlatie toe ⁽⁷⁰⁾, maar deze formules zijn zeer gecompliceerd. Hier biedt de homogeniteitsindex, die steeds dezelfde gedaante blijft behouden, ongeacht het aantal reactiemogelijkheden, grote voordelen. Wel neemt met het aantal reactiemogelijkheden de hoeveelheid rekenwerk, zij het ook niet geheel proportioneel, toe; maar men kan een eenvoudig rekenschema opstellen, dat door een statistisch niet geschoolde hulpkracht, zonder veel kans op fouten, kan worden toegepast. Het ligt daarom voor de hand, deze index ten behoeve van ons onderzoek te gebruiken.

Het rekenschema ten behoeve van de homogeniteitsanalyse heeft de volgende gedaante. Gebruikt wordt een formulier (afb. 5.1) met vijf kolommen voor de vijf verschillende itemscores, die van links naar rechts in opklimmende volgorde worden genoteerd.

Aan de linkerzijde van het formulier staan van boven naar beneden in opklimmende volgorde de schaalscores. Het gegeven over iedere proefpersoon wordt genoteerd in de kolom van zijn itemscore, op de rij van zijn schaalscore verminderd met de itemscore. Als dus b.v. zijn totale schaalscore 18 bedraagt en de itemscore 3, wordt dit in de vierde kolom, op rij 15 genoteerd („geturfd”). Nadat aldus voor één item de scores van alle proefpersonen zijn getabelleerd, worden voor de 2e t/m 5e kolom de cumulatieve frequenties van boven naar beneden bepaald. Zij worden telkens genoteerd op een rij, die één lager staat dan de rij waarop zij betrekking hebben. In de eerste kolom (nul-scores) geschiedt dit niet, hier worden alleen per rij de frequenties gebruikt.

Op een gegeven rij noemen we nu de frequenties in de opeenvolgende kolommen respectievelijk a t/m e en de daarachter genoteerde cumulatieve frequenties respectievelijk C_b t/m C_e. We berekenen dan de volgende grootheden:

HOMOGENITÄTSANALYSE

Item nr: 149

Gebied: VB

	a	C _a	b	C _b	c	C _c	d	C _d	e	C _e		A	B	C
0					//									
1	///		/		/	2						12	11	
2	///		//	1	/	3			//			34	43	
3	///		///	3		4	/			2		53	19	
4	///		/	7	///	4		1	/	2		58	22	
5	///		/	8		7	//	1		3		74	11	
6	///		/	9	///	7	//	3		3		235	65	
7	//		//	10	///	10		5	/	3		116	23	
8	/		/	12	///	13	//	5		4		91	17	
9	///		///	13		16		7		4		241	12	
10			///	16	/	16		7		4		119	4	
11				20	/	17	/	7	/	4		15	3	
12			/	20	/	18		8	/	5		44	3	
13	/			21	//	19		8	/	6		82	5	
14	/		/	21		21		8		7		93	1	
15			/	22		21		8		7		36		
16				23		21		8		7				
17				23		21	/	8		7		7		
18				23		21		9		7				
19				23	/	21		9		7		16		
20				23		22		9		7				
21		39		23		22		9		7				3668
Σ												1326	239	3668

850

$$H_{it} = 1 - \frac{2 \times 1326}{3668 - 239} = +.23$$

87

$$A = a (C_b + C_c + C_d + C_e) + b (C_o + C_d + C_e) + c (C_d + C_e) + d C_e$$

$$B = a (b + c + d + e) + b (c + d + e) + c (d + e) + d e$$

Deze beide grootheden worden over de rijen gesommeerd. Als verder T_a t/m T_o het totale aantal individuen voorstelt, dat in ieder der vijf kolommen is opgenomen, dan berekenen we vervolgens de grootheid

$$C = T_a (T_b + T_c + T_d + T_e) + T_b (T_c + T_d + T_e) + T_c (T_d + T_e) + T_d T_e$$

Uit de in figuur 5.1 afgebeelde rekenstaat blijkt gemakkelijk, dat ΣA het aantal heterogene onderscheidingen voorstelt, ΣB het aantal gevallen van „geen onderscheiding”, terwijl C gelijk is aan het totale aantal onderling vergeleken gevallen. Aangezien $H_o + H_e = C - \Sigma B$ kan de homogeniteitsindex thans worden bepaald met behulp van de formule

$$H_{it} = 1 - 2 \frac{\Sigma A}{C - \Sigma B}$$

Ten behoeve van de berekening der homogeniteitsindices is een steekproef van 100 personen gebruikt. Deze was verkregen door aselechte trekking uit het ten tijde van deze fase van het onderzoek ter beschikking staande materiaal, dat oorspronkelijk was verzameld met het oog op de constructie van de niveauschalen. In de steekproef komen dus voor: leerlingen van het v.h.m.o., m.u.l.o., lager technisch onderwijs en nijverheids-onderwijs voor meisjes. Allen waren leerlingen van de eerste klassen van scholen voor voortgezet onderwijs en nog slechts zeer korte tijd op het betreffende schooltype. Tabel 5.1. geeft een overzicht van de homogeniteitsindices der 162 items, gerangschikt naar de schalen, waartoe zij volgens IRLE moeten worden gerekend. De letteraanduiding van de gebieden en de nummeraanduiding van de items is dezelfde, die in hoofdstuk III werd gebruikt.

Vergelijken we de mediale waarden uit deze tabel met de mediale waarden der door IRLE bepaalde tetrachorische correlatiecoëfficiënten (zie paragraaf 2.3), dan valt op, dat onze uitkomsten over de gehele lijn aanzienlijk lager liggen. Dit is niet verwonderlijk, gezien de aard van de door ons gebruikte index. Immers, ook van KENDALL's *tau* is bekend, dat deze in de meeste gevallen aanzienlijk lager ligt, dan andere, verwante maatstaven. Volgens MC NEMAR (⁹⁷, pag. 210) is de waarde van *tau* in sommige gevallen slechts ongeveer twee-derde van de op hetzelfde materiaal bepaalde waarde van SPEARMAN's *rho*. Als we dit in het oog houden, kan de paralleliteit van onze uitkomsten met die van IRLE zeer bevredigend worden geacht. De meeste items, die voor een oudere leeftijdsgroep zijn ontworpen, bezitten ook reeds bij twaalfjarigen eenzelfde „Aufforderungscharakter”. Uitgaande van de verwantschap met KENDALL's *tau* kan men een significantiegrens voor de homogeniteitsindex vaststellen. Bij een steekproefomvang van 100 personen blijkt een *tau*-waarde van .18 reeds

TABEL 5.1. Homogeniteitsindices der items t.o.v. de bijbehorende schalen

TH		AV		TN		VB		AA	
item	H _{it}	item	H _{it}	item	H _{it}	item	H _{it}	item	H _{it}
1	.74	2	.44	3	.69	10	.20	12	.28
4	.81	5	.29	6	.60	13	.48	15	.56
7	.71	8	.49	9	.36	16	.23	18	.57
30	.75	29	.50	28	.69	38	— .27	39	.46
33	.80	32	.28	31	.62	41	.39	42	.58
36	.64	35	— .21	34	.68	44	.46	45	.44
57	.64	55	.44	56	.72	65	.27	64	.49
60	.70	58	.62	59	.57	68	.09	67	.51
63	— .04	61	.50	62	.58	71	.40	70	— .02
82	.74	83	.62	84	.55	91	.48	93	.37
85	.84	86	.05	87	.40	94	.41	96	.46
88	.82	89	.64	90	.57	97	— .04	99	.51
111	.71	110	.68	109	.81	119	.26	120	.33
114	.87	113	.58	112	.70	122	.08	123	.58
117	.79	116	.64	115	.78	125	.44	126	.53
138	.90	136	.59	137	.71	146	.28	145	.51
141	.61	139	.41	140	.86	149	.23	148	.35
144	.70	142	.56	143	.63	152	.00	151	.51
Mediaan .74		.50		.68		.27		.50	

TABEL 5.1. (vervolg)

HA		AD		LG		SO	
item	H _{it}	item	H _{it}	item	H _{it}	item	H _{it}
11	.09	19	.62	21	.36	20	.68
14	.38	22	.68	24	.50	23	.80
17	.41	25	.39	27	.41	26	.63
37	.32	48	.59	46	.60	47	.65
40	.28	51	.67	49	.58	50	.46
43	.31	54	.42	52	.49	53	— .21
66	.18	75	.41	74	.72	73	.64
69	.20	78	.26	77	.47	76	.54
72	.48	81	.60	80	.40	79	.66
92	.22	100	.44	102	.70	101	.80
95	.26	103	.38	105	.60	104	.62
98	.26	106	.39	108	.70	107	.69
118	.39	129	.57	127	.31	137	.86
121	.31	132	.33	130	.49	140	.85
124	.29	135	.55	133	.62	142	.76
147	.22	156	.52	155	.49	154	.84
150	.32	159	.32	158	.47	157	.75
153	.20	162	.48	161	.52	160	.80
Mediaan .29		.46		.50		.69	

op het 1%-niveau significant van nul te verschillen. Deze waarde kunnen we voor ons materiaal ook aanhouden. De zeer grote meerderheid der items heeft dan een homogeniteitsindex, die significant van nul verschilt en in vrijwel alle gevallen ligt het verschil in de gewenste richting (positieve waarde van H_{11}).

De aard van de uitzonderingen kan nog worden nagegaan: In de eerste schaal, technische handarbeid, heeft item 63: „stoelen vervaardigen” een nulcorrelatie. Deze past waarschijnlijk slecht bij de overwegend metaal-technische inslag van de overige items.

In de tweede schaal, ambachtelijke vormgeving, heeft item 86: „landschappen fotograferen” een nulcorrelatie. Men kan veronderstellen, dat dit komt, omdat ook onder degenen, die overigens weinig affiniteit tot deze schaal hebben, een voorkeur voor fotograferen veelvuldig voorkomt. Item 35: „landkaarten tekenen” heeft een significant negatieve correlatie. Aangezien deze item ook in de M-schaal een rol speelt, waar hij tezamen met natuurwetenschappelijke items voorkomt, ligt het voor de hand, dat deze item eerder tot de TN-schaal zal behoren.

In de schaal voedselbereiding, die ook voor IRLC vele problemen heeft opgeleverd, vallen vijf items uit. Nulcorrelaties hebben 68: „bier brouwen”, 97: „ossens aan het spit braden”, 122: „slachtvee op ziekte keuren” en 152: „scheikundige onderzoeken doen over levensmiddelen”. Een significant negatieve correlatie vertoont 38: „in een zuivelfabriek melk bacteriologisch onderzoeken”. Het ligt voor de hand, dat de beide laatstgenoemde items, met hun sterk natuurwetenschappelijke inslag, eerder in de schaal TN thuis zullen horen. Van item 122 valt de verwantschap met de schaal moeilijk in te zien, wellicht hoort een dergelijke item eerder bij de agrarische arbeid thuis. De twee eerstgenoemde items spreken vermoedelijk te weinig aan, zij worden door kinderen niet au sérieux genomen.

In de agrarische arbeid valt item 70: „op de haringvangst gaan” uit, hetgeen nauwelijks verrassend mag worden genoemd.

In de commerciële schaal past item 11: „de kaartverkoop in een circus organiseren”, niet. Ook dit is een enigszins buitenissige activiteit, waarbij bovendien het woord „circus” een verwarrende werking kan hebben.

Tenslotte heeft in de laatste schaal, sociaal werk en opvoeding, item 53: „loongeschillen beslechten” een significant negatieve correlatie. Ook dit is achteraf gemakkelijk begrijpelijk. De betreffende activiteit ligt niet in het vlak van de persoonlijke bemoeienis met mensen; zij draagt bovendien, mede door de vrij moeilijke formulering, een veel minder concreet karakter.

In totaal blijken dus 11 items niet in hun schaal te passen, d.w.z. ca. 7% van het totaal. Dit aantal is veel te gering, om een totale revisie van de test te rechtvaardigen, die zoals eerder uiteengezet, noodzakelijk zou zijn, om deze items door andere te vervangen.

Toch is ten aanzien van twee schalen de mogelijkheid van revisie overwogen: voedselbereiding en handel. In beide gevallen ligt namelijk de mediale waarde van de homogeniteitsindices veel lager dan bij de overige schalen. Deze bedraagt respectievelijk .27 en .29, terwijl de in grootte daaropvolgende mediale waarde .46 is. Nu verschillen de eerstgenoemde waarden wel significant van nul, maar het ligt voor de hand, dat men bij voorkeur in de schaal items zal opnemen, die wat hogere correlaties met het totaal vertonen.

In het geval van de schaal voedselbereiding is uiteindelijk van een revisie afgezien. Niet alleen op grond van de bekende praktische bezwaren tegen een herziening, maar ook omdat reeds bij de voorarbeid van IRLE gebleken is, dat het uitermate moeilijk is, geschikte items voor deze schaal te vinden, zo zelfs, dat IRLE er toe is overgegaan, twee items tweemaal op te nemen.

Ten aanzien van de tweede schaal lag de situatie enigszins anders. Bij bestudering van de items blijkt, dat in verschillende andere schalen activiteiten zijn opgenomen, die wellicht een commercieel aspect hebben. Het zou denkbaar zijn, dat met behulp van dergelijke items, tezamen met een aantal uit de bestaande HA-schaal, een nieuwe schaal zou kunnen worden geconstrueerd, met over het geheel meer homogene items. Door deze weg te volgen, vermijdt men ook de praktische bezwaren, die anders aan een schaal-revisie verbonden zijn, omdat nu de test als geheel in tact wordt gelaten. Het hierop betrekking hebbende onderzoek is in de volgende paragraaf gerapporteerd.

5.3. *Revisie van de HA-schaal*

Uitgangspunt voor de reconstructie is geweest, dat in de test verschillende items voorkomen, die niet in de HA-schaal zijn opgenomen en toch bij mentale analyse een commercieel aspect bezitten. Dit geldt b.v. voor: „een wijnhandel leiden” en „rapporten maken over de omzet van goederen”. Eventueel zouden ook: „de klantenadministratie van een groothandel bijhouden” of „een bijenteelt beginnen” een commercieel appèl kunnen uitoefenen.

Onderzocht moet worden, of dergelijke items homogeen zijn ten opzichte van de HA-schaal. Doordat de schaal in de oorspronkelijke vorm een aanzienlijk aantal items met betrekkelijk geringe homogeniteit bezit, bestaat het gevaar, dat het eventueel aanwezige commerciële aspect van de nieuw toe te voegen items voor een belangrijk gedeelte wordt gecamoufleerd door de invloed van de items, die we liever zouden willen elimineren. Daarom is eerst een zuivering van de bestaande HA-schaal uitgevoerd.

Dit geschiedde, door uit te gaan van de item, met de hoogste correlatie met de oude schaal, nr. 72 „koelkasten verkopen”. Mede gezien de eenvoudige aard van deze item werd geoordeeld, dat het begrip „commerciële

belangstelling", indien het op twaalfjarige leeftijd reeds reden van bestaan heeft, in ieder geval wel in deze item vertegenwoordigd zou zijn. Uit het beschikbare testmateriaal werd nu een steekproef geconstrueerd van 100 personen en wel 25 met een itemscore 0 op de item „koelkasten verkopen", 25 met itemscore 1, 25 met itemscore 2 en 25 met itemscore 3 of 4. Het aantal gevallen met een zeer hoge itemscore was in het destijds beschikbare materiaal te gering, om afzonderlijke groepen voor de itemscores 3 en 4 te vormen. De verdelingen van de itemscores van alle andere items uit de HA-schaal en ook die van een aantal andere items, waarvan werd verondersteld, dat zij voor opname in een nieuwe HA-schaal in aanmerking zouden kunnen komen, werden nu, met de verdeling van de itemscores op de kern-item, in een 5×4 tabel getabelleerd. Behalve de items uit de oude HA-schaal zijn op deze wijze onderzocht de items nr. 2, 8, 16, 22, 54, 91, 100, 106, 125, 149, 151 en 159. De items waarvan uit de tabel bleek, dat zij positief met het kernitem correleerden (hetgeen beoordeeld werd op de ligging der medianen in de vier rijen van de itemscores van item nr. 72), werden opgenomen in een voorlopige nieuwe HA-schaal. Ten opzichte van deze schaal zijn vervolgens de homogeniteitsindices van de items berekend. Deze berekening geschiedde met behulp van dezelfde steekproef, die ook voor de andere homogeniteitsberekeningen was gebruikt. Hierbij werden de volgende waarden voor H_{it} gevonden (tabel 5.2).

TABEL 5.2. Homogeniteit der items ten opzichte van de nieuwe HA-schaal

item	H_{it}	item	H_{it}	item	H_{it}	item	H_{it}
14	.62	43	.36	95	.10	125	.26
16	.39	54	.49	98	.48	150	.50
17	.57	69	.07	100	.60	153	.67
22	.38	72	.62	106	.43	159	.58
40	.41	92	.30	124	.19		

Uit het overzicht blijkt, dat de homogeniteit van de items met de schaal nu zeer aanzienlijk verbeterd is, de mediale waarde bedraagt thans .43 hetgeen past bij de waarden voor een aantal andere schalen. In de tabel treffen we nog enkele items met lage H_{it} aan, de nummers 69, 95 en 124. Deze zijn niet in de definitieve schaal opgenomen. Het ligt voor de hand, dat de homogeniteitsindices der resterende items enigszins zullen worden beïnvloed door deze eliminatie. Aangezien echter het effect vermoedelijk van geringe betekenis zal zijn en dus praktisch kan worden verwaarloosd, is de berekening van de definitieve waarden niet uitgevoerd. Uit de thans geconstrueerde schaal blijkt, dat een voldoende homogeen begrip „commerciële belangstelling" reeds bij degenen die op het punt staan de lagere school te verlaten, kan worden gehanteerd. De schaal is daarom in de

nieuwe vorm verder gebruikt, de later te rapporteren ijkingsgegevens hebben hierop betrekking.

5.4. Intercorrelaties

Het ligt voor de hand, dat de toepassing van een groot aantal schalen in een interessentest alleen zinvol is, indien deze schalen aspecten van de belangstelling toetsen, die in voldoende mate onderling onafhankelijk zijn. Enige correlatie zullen we in vele gevallen wel moeten verwachten, omdat verschillende arbeidsgebieden een beroep kunnen doen op verwante kwaliteiten. Omgekeerd kan ook een negatieve correlatie optreden ten gevolge van het feit, dat verschillende interesses in meerdere of mindere mate strijdig zijn.

De door IRLE bij zijn onderzoek van gemiddeld 16- à 17-jarigen gevonden intercorrelaties, hebben we reeds in paragraaf 2.3 gerapporteerd. Het is wel interessant, na te gaan, of het beeld bij onze, gemiddeld ca. 4 jaar jongere proefpersonen daarvan aanzienlijk afwijkt. Teneinde een juist beeld te krijgen, is het in dit geval van groot belang, een steekproef te gebruiken, die representatief is voor de verschillende tendenties in de bevolking. Wij hebben daarom een groep van 100 leerlingen van zesde klassen van lagere scholen samengesteld, die representatief beoogde te zijn voor de aantallen leerlingen, die naar de verschillende vormen van voortgezet onderwijs gaan.

In deze gestratificeerde steekproef werden opgenomen: 15 kinderen met bestemming v.h.m.o., 35 leerlingen met bestemming m.u.l.o. en 50 kinderen, die of lager voortgezet onderwijs zouden gaan volgen of geen verder dagonderwijs zouden gaan bezoeken. Deze keuze van de proefpersonen binnen de strata geschiedde aselekt uit bij ons aanwezig testmateriaal. Tabel 5.3 geeft de intercorrelaties voor het totale materiaal (50 jongens en 50 meisjes).

TABEL 5.3. Intercorrelaties BIT-Schalen

	AV	TN	VB	AA	HA	AD	LG	SO	L	M
TH	—,509	.634	—,329	.166	.302	.146	—,423	—,591	—,070	—,145
AV		—,581	.194	—,107	—,377	—,432	.199	.294	.310	—,095
TN			—,427	.022	.166	.197	.080	—,535	—,444	.438
VB				—,190	.184	.006	—,227	.241	.413	—,302
AA					—,151	—,058	—,271	—,392	—,083	—,296
HA						.678	—,279	—,384	—,511	—,240
AD							—,208	—,321	—,158	—,074
LG								.001	—,018	.734
SO									—,008	—,084
L										—,381

Bij vergelijking met de door IRLE gevonden waarden blijken, naast een groot aantal punten van overeenstemming, ook enkele verschillpunten.

Deze hebben voornamelijk betrekking op de gebieden HA en AD. Vermoedelijk zullen we dit moeten opvatten als een aanwijzing, dat de attitude ten opzichte van commercieel en administratief werk bij het gemiddelde kind op deze leeftijd nog niet zo duidelijk tot ontwikkeling is gekomen, als de instelling ten opzichte van verschillende andere delen van de beroepenwereld. Bij de beoordeling van de correlaties van de schalen M en L met de overige, dient te worden bedacht, dat hierbij tendele de score gebaseerd kan zijn op dezelfde items, zodat de correlatie niet op dezelfde wijze kan worden geïnterpreteerd. Laten we deze beide schalen daarom buiten beschouwing, dan kan van de overige worden gezegd, dat zij inderdaad over het geheel zodanig onderling onafhankelijk zijn, dat het gebruik in een belangstellingstest verantwoord is.

Deze conclusie kan worden bevestigd, door op de matrix der intercorrelaties van de eerste negen schalen een factor-analyse uit te voeren. Zelfs na extractie van vier factoren blijken nog steeds aanzienlijke restcorrelaties aanwezig te zijn. Ongeveer de helft van alle restcorrelaties heeft dan nog een absolute waarde tussen .050 en .150. Bovendien zijn de communaliteiten dan van dien aard, dat in zes van de negen gevallen de gemeenschappelijke factoren niet meer dan ten hoogste de helft tot de totale variantie bijdragen; dit wijst er op, dat er nog een aantal speciale factoren van niet te verwaarlozen betekenis aanwezig zijn. Tezamen genomen volgt hieruit de conclusie, dat een poging de beschrijving van de beroepeninteressen met behulp van een veel kleiner aantal schalen uit te voeren, niet verantwoord zou zijn.

5.5. *Betrouwbaarheid*

Naast de „homogeniteit” is ook de betrouwbaarheid van de test een belangrijk kenmerk. Traditioneel verlangt men, dat een test of testschaal een hoge betrouwbaarheidscoëfficiënt heeft, indien een toepassing voor beoordeling van individuen wordt beoogd. Hoe hoog de eisen zijn, die men in dat geval stelt, hangt af van het subjectief oordeel van de testgebruiker. De eerste onderzoekers op het terrein van testtheorie hebben wel eens een waarde van ca .95 als een minimum beschouwd (⁷³). Maar aangezien de ervaring leert dat tests met dergelijke betrouwbaarheden zeldzaam zijn, heeft men later wel enige concessies gedaan. Tegenwoordig wordt door velen een waarde van ca .80 à .85 als een aanvaardbare grootte beschouwd.

Er bestaan verschillende methoden, om de grootte van de betrouwbaarheidscoëfficiënt te bepalen: hertesten, gebruik van parallelle tests, splitsing van de test in gelijke delen en de methode van KUDER-RICHARDSON, die in wezen gebaseerd is op een variantie-analyse van de test-items. Van deze werkwijzen is de eerste weinig in tel, in verband met de vele foutenbronnen. De tweede komt in ons geval niet in aanmerking, omdat geen

parallel-test ter beschikking staat. De derde en de vierde methode kunnen beide worden toegepast.

Een keuze is hier moeilijk, omdat deze werkwijzen verwante merites en tekorten hebben.

Wij hebben daarom de bepaling van de betrouwbaarheidscoëfficiënten volgens beide methoden doorgevoerd. Daartoe was te meer aanleiding, omdat van de onderlinge verhouding van de uitkomsten van een betrouwbaarheidsbepaling volgens deze twee werkwijzen nog maar weinig met zekerheid bekend is. Het onderzoek biedt dus tevens een mogelijkheid, hierover informatie te verzamelen.

Bij de „split-halves” methode wordt de test verdeeld in twee helften, die gelijkwaardig kunnen worden geacht. De som van de scores op de twee testhelften denkt men gecorreleerd met de som van de scores op de twee overeenkomstige helften van een hypothetische paralleltest. Men veronderstelt nu, dat de covariantie van alle combinaties van twee testhelften dezelfde is. Op grond hiervan kan men gemakkelijk aantonen, dat de betrouwbaarheid r_{tt} gelijk is aan:

$$r_{tt} = \frac{4 r_{12} s_1 s_2}{s_1^2 + s_2^2 + 2 r_{12} s_1 s_2}$$

In deze formule is r_{12} de correlatie tussen de beide helften van de werkelijk bestaande test, terwijl s_1 en s_2 de standaarddeviaties van de scores op de eerste, respectievelijk tweede testhelft zijn. Deze drie waarden kunnen experimenteel worden bepaald. Onder de plausibele veronderstelling, dat s_1 en s_2 bij benadering aan elkaar gelijk zijn, kan deze formule worden vereenvoudigd tot de bekende formule van SPEARMAN-BROWN

$$r_{tt} = \frac{2 r_{12}}{1 + r_{12}}$$

Indien de variantie van de beide testhelften niet gelijk is, geeft de tweede formule een wat hogere waarde voor r_{tt} dan de eerste. De verschillen zijn echter van geringe praktische betekenis (⁹¹, hoofdstuk 15). We hebben daarom de eenvoudige formule van SPEARMAN-BROWN toegepast.

De bepaling van de betrouwbaarheid volgens KUDER-RICHARDSON is uiteraard gebruikt bij het onderzoek van de KUDER Preference Record, terwijl ook IRLE bij het betrouwbaarheidsonderzoek er gebruik van maakt. De beste uitkomsten levert de zogenaamde „formule 20” (vergelijk ⁵⁸, hoofdstuk 16).

Deze formule, die op zeer verschillende wijze kan worden afgeleid (^{40a}, ⁵⁸, ⁶⁵, ⁹¹, ^{131a}), luidt:

$$r_{tt} = \frac{N}{N-1} \left(\frac{s_t^2 - \sum p_i q_i}{s_t^2} \right)$$

waarin N het aantal items in de test, p_i het gedeelte van de proefgroep, dat

item i verkiest (c.q. goed beantwoordt) en s_i^2 de variantie van de totale test is; $q_i = 1 - p_i$.

Onder de veronderstelling, dat p_i constant is, kan deze formule worden vereenvoudigd tot formule 21, die, enigszins omgewerkt, luidt:

$$r_{tt} = \frac{Ns_t^2 + M(M-N)}{s_t^2(N-1)}$$

waarin M de gemiddelde test- of schaalscore is.

LORD⁽⁹⁴⁾ heeft echter aangetoond, dat formule 21 ook kan worden verkregen zonder de veronderstelling, dat p_i constant is. In zijn afleiding wordt alleen gebruik gemaakt van de veronderstellingen, dat we

- r_{tt} opvatten als een correlatie tussen aselekt parallele tests, d.w.z. tests, die gedacht worden te bestaan uit items, die aselekt uit één grote verzameling items zijn getrokken;
- gebruik mogen maken van een „efficient unbiased” schatting van de standaard meetfout, in plaats van de onbekende ware standaardfout te gebruiken;
- de formule betrekken op een groot aantal onderzochte individuen.

Door deze afleiding worden de toepassingsmogelijkheden van formule 21 aanzienlijk verruimd. Onderzoek⁽⁹¹⁾, pag. 592 heeft geleerd, dat formule 21 iets lagere waarden voor de betrouwbaarheid geeft dan formule 20. Het verschil is echter bij een test van enige omvang niet zeer groot, het ligt dan in de grootte-orde van .050. Op grond van deze overwegingen hebben wij formule 21 gekozen.

Voor de bepaling van de betrouwbaarheid der schalen is gebruikt een steekproef van ruim 1000 leerlingen van zesde klassen. Deze steekproef, die ook voor de landelijke ijking van de test is gebruikt, zal in het volgende hoofdstuk nader worden beschreven. Zij is naar de sexe verdeeld in twee subgroepen, bestaande uit 531 jongens en 487 meisjes.

Ten behoeve van de „split-halves” berekening is de test in twee gelijke helften verdeeld, door de score op het eerste blad van het testformulier te correleren met de score op het tweede blad. Zoals bekend zijn de twee helften van de test geheel op dezelfde wijze opgebouwd, zodat verondersteld mocht worden, dat aan de basis-veronderstelling van de „split-halves”-methode voldaan werd. De daarbij gevonden correlatiecoëfficiënten zijn volgens SPEARMAN-BROWN gecorrigeerd.

De door ons toegepaste vorm van de „split-halves”-methode, is niet de meest gebruikelijke. Gewoonlijk wordt de even - oneven nummering van de items als splitsingsprincipe gekozen. Aangezien echter de items uit de schalen aselekt over de beide bladzijden verdeeld zijn, kunnen voorkant en achterkant als twee paralleltests worden beschouwd, even goed als dat met de even en oneven genummerde items het geval is. Bij contrôle is gebleken, dat de variantie op vrijwel alle schalen op beide pagina's van de test

nagenoeg dezelfde is. Daarmee is ook het gebruik van de formule van SPEARMAN-BROWN gerechtvaardigd.

In tabel 5.4 zijn de resultaten van de betrouwbaarheidsbepaling volgens de beide beschreven methoden weergegeven.

TABEL 5.4. Betrouwbaarheidscoëfficiënten volgens „split-halves” en KUDER-RICHARDSON

	jongens		meisjes	
	spl.	K.R.	spl.	K.R.
TH	.798	.936	.727	.809
AV	.858	.875	.769	.826
TN	.854	.918	.777	.812
VB	.748	.801	.644	.640
AA	.906	.944	.841	.880
HA	.660	.761	.690	.772
AD	.858	.907	.778	.837
LG	.851	.927	.787	.863
SO	.882	.933	.867	.929
L	.512	.706	.255	.495
M	.662	.870	.613	.752

Uit deze tabel blijken enkele consistente kenmerken. De betrouwbaarheid van de scores is bij de jongens vrijwel over de gehele lijn hoger dan bij de meisjes. Dit wijst er op, dat de items van de schalen over het geheel wellicht te eenzijdig in de manlijke belangstellingsrichtingen zijn gekozen. In dat geval immers zullen de antwoorden van de meisjes, die door het principe van de gedwongen keuze toch even frequent worden gegeven als die van de jongens, in sterker mate worden bepaald door de nevenaspecten van de items. Ten opzicht van de schalen betekent dit, dat de beantwoording bij de meisjes meer een toevalskarakter krijgt.

In de tweede plaats zijn de coëfficiënten volgens KUDER-RICHARDSON vrijwel over de gehele lijn aanzienlijk hoger dan die volgens de „split-halves”-methode. Dit is des te opvallender, omdat de gebruikte formules eerder een effect in tegengestelde richting zouden doen verwachten. De verwachting is immers, dat de formule van SPEARMAN-BROWN iets te hoge en formule 21 iets te lage uitkomsten zou geven.

Aanvankelijk leek de veronderstelling voor de hand te liggen, dat dit veroorzaakt zou worden door een vermoeidheidseffect, waardoor de beantwoording van de tweede pagina sterker dan die van de eerste, door toevallige factoren beïnvloed zou kunnen worden. Ter verificatie van deze veronderstelling zijn voor enkele schalen nog de even-oneven betrouwbaarheden berekend. Deze bleken echter, met één uitzondering, niet aanzienlijk of op gerichte wijze van de door ons berekende coëfficiënten te verschillen. De verklaring van de grote verschillen is daarom voorschijns niet duidelijk, maar in ieder geval houden deze cijfers de waarschuwing in, dat

betrouwbaarheidscoëfficiënten, die met behulp van verschillende methoden zijn verkregen, niet zonder meer als equivalent moeten worden beschouwd.

De coëfficiënten volgens KUDER-RICHARDSON stemmen naar hoogte goed overeen met de waarden, die IRLE heeft gevonden (vergelijk paragraaf 2.3). De test is dus, wat betrouwbaarheid betreft, bij 12-jarigen evenzeer bruikbaar als bij ouderen.

De twee zorgenkinderen van de test blijken, vooral bij de meisjes, de VB- en de L-schaal te zijn. Weliswaar is de betrouwbaarheid van laatstgenoemde schaal beter dan de in tabel 5.4 opgegeven waarde suggereert. Bij de berekening van de even-oneven betrouwbaarheidscoëfficiënten werd namelijk voor deze schaal, bij de meisjes een veel hogere waarde gevonden, namelijk .786 (bij de jongens .563). Vermoedelijk wordt dit grote verschil veroorzaakt door het feit, dat de items van de L-schaal zeer onregelmatig over de test verdeeld zijn, zodat in dit geval de eerste en tweede pagina ten onrechte als paralleltests zijn behandeld.

Hoewel de opgegeven waarden nog steeds van dien aard zijn, dat toepassing van de scores in individuele gevallen met veel voorzichtigheid moet geschieden, kan toch op deze plaats wel reeds worden meegedeeld, dat bij het later te bespreken onderzoek naar de synchrone validiteit van de test is gebleken, dat zowel de VB- als de L-schaal in verschillende situaties waardevolle indicatoren zijn gebleken.

Daarnaast geldt natuurlijk, zoals onder anderen KUDER⁽⁸¹⁾, pag. 6) onlangs concreet heeft gedemonstreerd, dat ook tests met een zeer lage betrouwbaarheid praktisch kunnen worden gebruikt, indien men de testgegevens verwerkt in een predictor-formule. In een dergelijk geval kan een combinatie van een aantal gegevens van geringe betrouwbaarheid zelfs een betere predictieve validiteit blijken te bezitten, dan de score op één even relevante test met een grotere betrouwbaarheid. Aangezien echter, zeker in de Nederlandse situatie, het gebruik van dergelijke predictor-formules, waarbij de individuele interpretatie wordt uitgeschakeld, nog niet aan de orde is, is deze opvatting over de waarde van tests met een lage betrouwbaarheid voorlopig alleen van theoretisch belang.

HOOFDSTUK VI.

DE IJ KING VAN DE B I T-SCHALEN

6.1. *De experimentele opzet*

De problematiek van de ijking van een test van het type van onze belangstellingstest biedt in theorie weinig complicaties. Indien men er in slaagt, de populatie voldoende scherp te definiëren, is het voornaamste probleem het trekken van een steekproef van voldoende representativiteit. Op dit vraagstuk zal daarom vrij uitvoerig worden ingegaan.

Wij willen de test bruikbaar maken voor onderzoek van leerlingen van de zesde klasse van lagere scholen in ons land. Daarmee kan de populatie, waaruit de steekproef moet worden getrokken adaequaat worden aangeduid: deze omvat alle leerlingen, zowel manlijke als vrouwelijke, uit alle zesde klassen van lagere scholen in Nederland. De steekproef moet daarvan zo goed mogelijk een afspiegeling zijn. Nadat de steekproef is gekozen, zal dus, voorafgaand aan de ijking, moeten worden onderzocht, of de steekproef inderdaad representatief is.

De laatste tijd zijn enkele malen landelijke ijkingen van psychologische tests ondernomen. Door leden van de „Researchwerkgroep Schoolpsychologie” is samengewerkt bij de ijking van een door Prof. KOUWER ontwikkelde vorm van een intelligentietest (6). Daartoe hebben psychologen en psychologische instituten in verschillende delen van het land tests afgenomen op scholen, die gekozen waren op grond van een stratificatieschema. Hierin werd rekening gehouden met de districten van het land, het welstandsniveau van de gezinnen en de gemeentegrootte. Bij de ijking van een door ons ontwikkelde test voor de mechanische leesvaardigheid (138) werd hetzelfde schema gevolgd. Ditmaal geschiedde het afnemen van de test door medewerkers van de afdeling Geestelijke Gezondheid van het Nederlands Instituut voor Praeventieve Geneeskunde, die op de, in overleg met de onderwijsinspecteurs gekozen scholen, onderzoek verrichtten.

In beide gevallen betrof de ijking tests, die een zekere mate van vaardigheid bij de proefleider veronderstellen. De B I T daarentegen is een test, die ook door een niet-psycholoog, die geen ervaring heeft in het afnemen van tests, zonder veel moeite kan worden toegepast. De instructie is eenvoudig en er kunnen zich tijdens het onderzoek slechts weinig complicaties voordoen. De in paragraaf 3.3 gegeven handleiding is op vrijwel alle testsituaties berekend. Op grond van deze overwegingen werd besloten, een proef te nemen met een ijking, die gebaseerd was op de mede-

werking van klasse-onderwijzers, die, zonder supervisie ter plaatse, de test afnamen. Scoring en verdere bewerking geschieden vanzelfsprekend centraal.

De voordelen van deze werkwijze liggen voor de hand: zij is economisch, maakt het mogelijk binnen kort tijdsbestek, over het gehele land verspreid, een groot aantal kinderen te onderzoeken. Alle onderzochten verkeren daardoor in hetzelfde onderwijsstadium, hetgeen aan de homogeniteit van het ijkingsmateriaal ten goede komt. Besloten werd, de ijking omstreeks het einde van het schooljaar uit te voeren.

De grootte van de ijkingsgroep werd vastgesteld op ca 1000 kinderen. De gegevens, op dit aantal betrekking hebbende, waren met de beschikbare apparatuur nog relatief gemakkelijk te hanteren en te bewerken; anderzijds is het aantal voldoende groot om de invloed van toevallige variaties tot geringe proporties terug te brengen. Gebleken is dat bij de gekozen steekproefgrootte de mediale standaardfout van het gemiddelde ruim .1 bedraagt, zodat de ligging van het gemiddelde der schalen als regel op minder dan een half punt nauwkeurig kan worden vastgesteld.

Gezien de aard van de test lag het voor de hand, gehele schoolklassen in het onderzoek te betrekken. Dit maakte het ook gemakkelijker, de medewerking van onderwijzend personeel te vragen. Op grond van deze gedachte werd de steekproef samengesteld door aselechte trekking van scholen, uit een adreslijst (³). De gegevens uit deze adreslijst zijn natuurlijk niet geheel up to date. De in de loop van enkele jaren optredende veranderingen zijn echter, gezien het grote aantal scholen, betrekkelijk weinig belangrijk, terwijl zij bovendien in alle delen van het land voorkomen en dus op de nationale verdeling van de leerlingen amper invloed hebben.

Iedere gewone lagere school, die in deze adreslijst voorkwam, kreeg een fictief nummer. Dit werd bepaald door het nummer van de bladzijde (drie cijfers) en het rangnummer van de school op de bladzijde (twee cijfers), de karakteristiek bestond dus uit een getal van vijf cijfers. Door trekking uit een tabel van aselechte getallen werden vervolgens getallen van vijf cijfers gevormd, waarbij een aantal van de begincijfers nul mocht zijn. Voor zover de zo geconstrueerde getallen corresponderen met scholen uit de adreslijst, werden deze in de steekproef opgenomen.

Over het aantal op te nemen scholen moest een schatting worden gemaakt. Besloten werd in eerste instantie zo veel scholen aan te schrijven, dat daarmee maximaal het onderzoek betrekking zou hebben op ca 1400 kinderen. Het werkelijke aantal proefpersonen was uiteraard afhankelijk van de mate, waarin de scholen bereid zouden zijn, hun medewerking te verlenen. Hierover was zelfs bij grove benadering geen realistische schatting mogelijk. In de betreffende adreslijst stond bij iedere school het totale leerlingental uit 1955 vermeld, het aantal zesde klassers kan op iets minder dan één-zesde van dit aantal worden gesteld. Besloten werd daarom, met het trekken van de steekproef voort te gaan, tot de som van de totale

leerlingenaantallen tussen 8500 en 9000 lag. In totaal werden op deze wijze 43 scholen met, volgens de adreslijst, 8810 leerlingen gekozen.

Aan de hoofden van al deze scholen werd een brief gezonden, waarin in het kort een verzoek om medewerking werd gedaan. Gezegd werd, dat deze brief aan 43 scholen werd verzonden en dat het deelnemen van alle scholen van groot belang was, in verband met de poging een landelijk representatieve ijking door te voeren. Ter toelichting werd verwezen naar een gestencild dossier, dat bij de brief was ingesloten. Het bijgevoegde dossier bevatte in de eerste plaats een korte beschrijving van het onderzoek, waarin o.a. de aard van de gevraagde medewerking werd gepreciseerd. Toegezegd werd, dat de school op haar betrekking hebbende gegevens, na de uitwerking zou vernemen en dat deze gegevens niet aan derden zouden worden medegedeeld. Voorts was opgenomen een beschrijving van de test, met de testinstructie, waarbij een exemplaar van de test was gevoegd. Tenslotte bevatte het dossier een deelnameformulier, waarop kon worden ingevuld hoe groot het aantal leerlingen van de zesde klas was en wanneer men dacht, dat het onderzoek gereed zou kunnen zijn. Na ca. drie weken werd aan alle scholen, die nog niet geantwoord hadden, een rappel gezonden. Op deze wijze kon worden bereikt, dat van 34 scholen antwoord werd ontvangen. Hiervan deelden 8 scholen mede, dat zij geen medewerking konden verlenen. Enige plattelandsscholen omdat de zesde klasse geen of slechts enkele leerlingen bevatte, één school om interne redenen (verbouwing), terwijl een aantal geen reden opgaf. Het aantal proefpersonen dat daarmee kon worden bereikt, bedroeg 786, d.w.z. ca. 60% van het geschatte aantal leerlingen, dat in de zesde klassen der aangeschreven scholen zat. Uit de eerste binnengekomen zendingen testformulieren was inmiddels gebleken, dat rekening moest worden gehouden met ca. 5% onbruikbaar materiaal (mislukte tests, ontbreken van de aanvullende gegevens, e.d.). Daarom werd besloten, volgens de reeds beschreven methode, zoveel nieuwe scholen te trekken, dat op basis van een deelname van 60% een totaal van ca. 1100 proefpersonen zou worden bereikt. Het bleek nodig, daartoe 18 scholen nieuw aan te schrijven. In 9 gevallen werd bericht ontvangen, dat men tot medewerking bereid was. Aangezien daarmee het gestelde aantal proefpersonen reeds bereikt bleek, is bij de tweede groep niet gerappelleerd.

Aan de deelnemende scholen werd, behalve het benodigde aantal testformulieren en een stencil met aanwijzingen voor de proefleider, een lijst gezonden, waarop een aantal aanvullende gegevens over de kinderen werd gevraagd. De vragen hadden betrekking op naam, voornamen, geboortedatum en sexe van de leerlingen, op de klassen, die eventueel waren overgeslagen of gedoubleerd, op het beroep van de vader, op de welstand van het gezin en op de bestemming van het kind na het verlaten van de zesde klas. Bovendien was een kolom voor opmerkingen aanwezig. In de aanwijzingen, die op de vragenlijst betrekking hadden, werd o.a. ten

behoefte van de schatting van het welstandsniveau een schaal met zeven punten, die in een omschrijving gekarakteriseerd werden, gegeven. Verzocht werd voorts, bij het vermelden van de bestemming, rekening te houden met de mogelijkheid, dat het kind een overgangsfase zou doormaken en in dat geval ook de volgende bestemming te vermelden (b.v. via 7e klas naar lyceum, via v.g.l.o. naar landbouwschool, e.d.).

De medewerking van de scholen is zeer vlot verlopen, zodat in ongeveer twee maanden tijds de formulieren, betrekking hebbende op 1070 kinderen, werden ontvangen. In twee gevallen bleek een misverstand te zijn ontstaan, waardoor de testformulieren onbruikbaar waren, in beide gevallen is in overleg met het hoofd van de school de test in de gehele klasse herhaald. Bij controle bleken 36 formulieren (3.4%) onbruikbaar, vrijwel steeds omdat het betreffende kind met de test niet klaar was gekomen, een enkele maal ook omdat slechts de helft van de test was gemaakt. Bovendien vielen voor de ijking uit, de gegevens betrekking hebbende op 16 leerlingen, die op kleine scholen in het zevende of achtste leerjaar waren geplaatst, maar die tezamen met de zesde klasse onderwijs ontvingen.

De formulieren werden gescoord op elf schalen: de gewijzigde HASchaal, de acht overige schalen van IRLE en de beide niveauschalen L en M. Deze scores werden per school en per kind geregistreerd.

Voor de verdere bewerking werd gebruik gemaakt van I.B.M.-Hollerith-apparatuur. Beschikbaar waren ponskaarten met 80 kolommen, waarvan er 4 voor de administratie van de Hollerith-afdeling gereserveerd moesten blijven. Het bleek mogelijk, alle gewenste gegevens op één kaart onder te brengen. De indeling van de ponskaarten werd de volgende:

- Kolom 1—4 gereserveerd voor Hollerith-afdeling
- 5 richting van de school (openbaar, Prot. Chr., R.K. jongens, meisjes of gemengd, Bijz. Neutr.)
- 6 nummer van de school
- 7—8 nummer van het kind (de ponsingen 00 t/m 49 werden gereserveerd voor meisjes; de ponsingen 50 t/m 99 voor jongens)
- 9 aard woonplaats
- 10 geboortjaar kind
- 11 geboortemaand kind
- 12 beroepsniveau vader
- 13 beroepensector vader
- 14 bestemming
- 15—16 schaal VB voorzijde
- 17—18 schaal VB achterzijde
- 19—20 schaal VB totaalscore

Van de andere tien schalen werden op analoge wijze eveneens telkens in zes kolommen drie scores vastgelegd.

Ter toelichting diene, dat door de wijze van registreren in kolom 7—8 geen afzonderlijke codering van de sexe behoefde te geschieden. Indien op een school meer dan 50 jongens of meisjes aan het onderzoek deelnamen, kreeg de betreffende school in kolom 6 twee nummers toegewezen. De nummering geschiedde per richting; aangezien meer dan 10 openbare scholen aan het onderzoek hadden deelgenomen, kreeg deze onderwijsvorm in kolom 5 twee nummers.

De aard van de woonplaats werd aangegeven door de onderscheiding: zeer grote stad, grote stad (> 100.000 inwoners), middelgrote stad (> 50.000 inwoners), kleine stad, geïurbaniseerd gebied, overgangsgebied, platteland. De provincie werd niet geregistreerd, omdat deze gemakkelijk op grond van de nummering van de scholen terug te vinden zou zijn.

Bij de registratie van het geboortjaar werden afzonderlijke nummers voor het eerste en het tweede halfjaar gebruikt, daardoor kon voor de geboortemaand met één kolom worden volstaan.

Het niveau van het beroep van de vader — c.q. het beroep dat deze het laatst had uitgeoefend — werd in een negenklassige schaal ingedeeld, die ontleend was aan TOBI en LUYCKX ⁽¹³⁰⁾. Beroepen, die niet in hun lijst voorkomen, werden, door vergelijking met verwante arbeid, zo goed mogelijk ingedeeld. Bij deze wijze van indelen is het terrein, waarop de vader werkzaam was, niet meer uit de gegevens te achterhalen. Daarom werd ter aanvulling in kolom 13 de sector van zijn beroep, volgens de indeling die door het C.B.S. wordt gebruikt, vastgelegd. Deze sector-indeling is: diensten, industrie en ambacht, handel, administratie, gezag, verkeer, vrije beroepen en onderwijs, diversen c.q. onbekend, geen arbeid.

Tenslotte werd de bestemming van het kind vastgelegd, ingedeeld naar v.h.m.o., m.u.l.o. en handelsdagonderwijs, n.o. voor jongens, n.o. voor meisjes, overig vakonderwijs, v.g.l.o., de praktijk in, diversen, bestemming onbekend, brugklasse. Behalve in het geval van overgang naar een brugklasse is, indien van de eerste bestemming bekend was, dat deze slechts als een kort intermediair bedoeld was, bij de registratie de verdere bestemming vastgelegd.

De op deze wijze geregistreerde gegevens omvatten meer dan voor de ijking van de test noodzakelijk is. Zij waren tevens bedoeld voor een aantal onderzoeken over de karakteristiek van de belangstelling bij leerlingen omstreeks het einde van de lagere school. Over deze onderzoeken zal in enkele volgende hoofdstukken worden gerapporteerd.

6.2. *De representativiteit van de steekproef*

De representativiteit van een steekproef kan worden onderzocht, door de karakteristiek van de proefpersonen in verschillende opzichten te vergelijken met de gegevens, die over de karakteristiek van de populatie bekend zijn. In hoe meer opzichten er overeenstemming is en hoe nauwer deze overeenstemming is, des te groter wordt de kans, dat ook de ijkings-

gegevens representatief zullen zijn. In dit onderzoek hebben wij de steekproef in zeven opzichten vergeleken met de populatie.

a. De regionale verdeling

In tabel 6.1 is de verdeling van de proefpersonen over de provincies weergegeven, met daarnaast de verdeling van de populatie. Aangezien geen recente gegevens over de verdeling van de schoolbevolking uit de zesde klassen der provincie ter beschikking stonden, zijn de C.B.S.-gegevens over de totale bevolking gebruikt ⁽³⁰⁾. In de tabel is voorts vermeld, uit welke plaatsen de proefpersonen afkomstig zijn.

TABEL 6.1. Verdeling van de steekproef over de provincies

Prov.	Aantal p.p.	Perc.	Popula- tie	Onderzoek-plaatsen
Gr.	18	1.7%	3.9%	Sappemeer
Fr.	68	6.6%	4.3%	Akkerwoude, Leeuwarden Sexbierum
Dr.	—	—	2.8%	Enschede, Hengelo
Ov.	78	7.5%	6.9%	Angerlo, Apeldoorn, Ede, Ewijk,
Geld.	116	11.2%	11.4%	Nijmegen (2×), Silvolde
Utr.	133	13.0%	5.9%	Bunnik, Kamerik, Zeist
N.H.	184	17.8%	16.6%	Amsterdam (3×), Bovenkerk, Enk- huizen
Z.H.	110	10.6%	22.4%	Den Haag (2×), Rotterdam, Vlaar- dingen, Wassenaar
Zeel.	25	2.4%	2.3%	Bruinisse
N.B.	247	23.8%	14.6%	Alphen, Breda, Made, Tilburg, Zeelst
Limb.	41	4.0%	8.5%	Nederweert
N.O.P.	14	1.3%	.3%	Rutten

De overeenstemming tussen de regionale verdeling van de steekproef en die van de populatie blijkt over het algemeen goed te zijn. Dit blijkt met name, als we de provincies samen nemen tot vier districten:

Noord (Gr., Fr., Dr., N.O.P.)	gevonden 9.6%	verwacht 11.3%
Oost (Ov., Geld.)	gevonden 18.7%	verwacht 18.3%
West (Utr., N.H., Z.H., Zeel.)	gevonden 43.8%	verwacht 47.2%
Zuid (N.B., Limb.)	gevonden 27.8%	verwacht 23.1%

b) Verdeling naar gemeente-grootte

Aangezien landelijke gegevens over de indeling, die in onze codering was gemaakt, niet op eenvoudige wijze ter beschikking staan, zijn de onderzoekplaatsen her-ingedeeld naar de gemeentegrootte. In tabel 6.2 wordt de verdeling van de populatie met die van de C.B.S.-gegevens vergeleken, waarbij weer dezelfde bron als in het voorgaande geval is gebruikt.

TABEL 6.2. Verdeling van de steekproef over de gemeenten

Inwonertal	Aantal ppn.	Percentage	Percentage populatie
minder dan 5000	141	13.6	13.9
5.000— 10.000	137	13.2	14.6
10.000— 20.000	90	8.7	15.9
20.000— 50.000	64	6.1	14.7
50.000—100.000	243	23.5	12.4
100.000 en meer	359	34.8	28.4

In de plattelandsgemeenten en de grote steden blijkt de overeenstemming dus zeer goed te zijn, de kleine steden zijn in te gering aantal vertegenwoordigd, ten gunste van de middelgrote steden. Bij het validatie-onderzoek zijn tussen deze beide categorieën geen significante verschillen gevonden, zodat deze afwijking op de representativiteit van de steekproef geen wezenlijke invloed zal hebben.

c. De sexe-verdeling

Onze steekproef bestaat uit 487 meisjes (47.8%) en 531 jongens (52.2%). Volgens vrij recente C.B.S.-gegevens ⁽³³⁾ bestond de bevolking van het zesde leerjaar uit 48.9% meisjes en 51.1% jongens. De overeenstemming is dus goed. Verschillen in dit opzicht zouden voor de ijking trouwens geen betekenis hebben gehad, omdat de gegevens van jongens en meisjes afzonderlijk worden behandeld.

d. De leeftijdsverdeling

Betrekkelijk recente gegevens over de bevolking van de zesde klasse kunnen worden ontleend aan een publicatie van het K.S.K.I. ⁽⁹⁾. Aangezien hierin echter geen gegevens over de 11-jarigen waren opgenomen, terwijl ook uit een andere bron over de leeftijdsverdeling in de lagere school ⁽³¹⁾, geen voldoende informatie kon worden geput, is deze groep buiten beschouwing gelaten. Daartoe is trouwens ook aanleiding, omdat de leeftijdsverdeling afhankelijk is van de datum, waarop de gegevens worden verzameld. De gegevens over de 12-jarigen en ouderen zijn omgerekend op 100% (tabel 6.3).

TABEL 6.3. Leeftijdsopbouw van de steekproef

	Meisjes		Jongens	
	Steekproef	Schoolbev.	Steekproef	Schoolbev.
12 jaar	68.1%	75.1%	69.1%	69.3%
13 jaar	28.2%	21.5%	27.3%	25.4%
14 jaar	3.7%	3.4%	3.6%	5.3%

In aanmerking genomen, dat het tijdstip van de leeftijdsbepaling invloed heeft op verdeling, kan de overeenstemming tussen steekproef en populatie zeer bevredigend worden genoemd. Het percentage oudere leerlingen is in ons materiaal iets groter dan in de populatie, hetgeen in overeenstemming is met het feit, dat ons onderzoek aan het einde van het schooljaar plaatsvond.

e. Beroep van de vader

De gegevens betreffende de grootste beroepensectoren zijn vergeleken met de uitkomsten van de beroepentelling van 1947. De vier grootste groepen omvatten de vaders van ruim 800 kinderen. Tabel 6.4 geeft een beeld van de zeer goede overeenstemming tussen steekproef en populatie.

TABEL 6.4. Beroepen van de vaders der kinderen uit de steekproef

	landbouw	industrie en ambacht	handel	verkeer
steekproef	16.6%	44.9%	11.3%	9.8%
beroepsbevolking	18.0%	44.7%	12.5%	10.0%

f. Bestemming van de leerlingen

Voor de vergelijking hebben wij de kinderen ingedeeld in vier grote groepen: v.h.m.o. (inclusief brugklassen, e.d.); m.u.l.o.; dag-n.o. en ander vakonderwijs; part-time onderwijs of geen onderwijs. De v.g.l.o.-leerlingen zijn bij de laatstgenoemde groep gerekend. Ter vergelijking zijn de door het C.B.S. gepubliceerde gegevens over 1955 gebruikt (³⁸). In aanmerking genomen, dat de toewijzing tot de verschillende categorieën vóór het einde van het schooljaar altijd enige onzekerheid impliceert, kan de overeenstemming van de in tabel 6.5 vermelde percentages opvallend goed worden genoemd.

TABEL 6.5. Bestemming van de leerlingen uit de steekproef

	v.h.m.o.	m.u.l.o.	dag-n.o.	part-time of geen onderwijs
<i>meisjes</i>				
steekproef	11.7%	33.7%	32.2%	22.3%
C.B.S.-gegevens	10.7%	32.7%	36.7%	19.9%
<i>jongens</i>				
steekproef	16.9%	28.8%	38.0%	16.2%
C.B.S.-gegevens	16.0%	31.4%	32.4%	20.2%
<i>totaal</i>				
steekproef	14.4%	31.1%	35.2%	19.1%
C.B.S.-gegevens	13.4%	32.0%	34.5%	20.1%

Een steeds groter percentage leerlingen volgt na het verlaten van de lagere school verder onderwijs. Men zou dus mogen verwachten, dat in ons materiaal de percentages in de drie eerste categorieën wat hoger zouden zijn dan in de vergelijkingsgegevens. Vermoedelijk zijn echter jongens met bestemming m.u.l.o. en meisjes met bestemming n.o. in wat te geringe getale in het materiaal vertegenwoordigd.

g. De leerlingen naar richting

Zuilsgewijs zijn de scholen in vier categorieën in te delen: openbaar, protestant-christelijk, rooms-katholiek en bijzonder neutraal. Tabel 6.6 geeft, gesplitst voor de sexen, een vergelijking tussen de gegevens uit de steekproef en die betreffende het gehele lager onderwijs. De verdeling van de leerlingen uit de zesde klassen is niet precies bekend, maar er is geen reden aan te nemen, dat deze aanmerkelijk van die van de totale schoolbevolking zou afwijken. De vergelijkingsmaatstaf heeft betrekking op de C.B.S.-gegevens ⁽³⁸⁾ van 1956.

TABEL 6.6. Verdeling van de leerlingen naar richting

	Openbaar	Prot. Chr.	R.K.	Bijz. Neutr.
<i>Meisjes</i>				
Steekproef	24.3%	4.6%	18.6%	.4%
Populatie	13.8%	13.2%	20.4%	.9%
<i>Jongens</i>				
Steekproef	24.0%	5.3%	22.2%	.6%
Populatie	14.9%	14.1%	21.6%	1.0%
<i>Totaal</i>				
Steekproef	48.2%	9.9%	40.9%	1.0%
Populatie	28.7%	27.3%	41.9%	1.9%

In verband met de wijze waarop de C.B.S.-gegevens zijn gerangschikt, zijn de percentages van de jongens en meisjes in de verschillende richtingen hier berekend op de totale populatie.

Uit de tabel blijkt, dat het R.K. en Bijzonder-Neutraal onderwijs in de juiste proporties in de steekproef vertegenwoordigd is. Leerlingen van het Prot. Chr. onderwijs komen daarentegen in veel te kleine getale voor, terwijl er relatief te veel leerlingen van het openbaar onderwijs in de steekproef voorkomen. Dit is een direct gevolg van de medewerking van de overeenkomstige scholen. Hier ligt het enige punt, waar de steekproef een enigszins ernstige tekortkoming vertoont. Daarom is onderzocht, of dit gecorrigeerd zou moeten worden, door de resultaten van de leerlingen van het openbaar en protestant-christelijk onderwijs gewogen in de ijking te verwerken. Bij het validatie-onderzoek is echter gebleken, dat er geen noemenswaardig verschil tussen het voorkeurspatroon van de leerlingen

van deze beide schooltypen bestaat (vergelijk hoofdstuk XII), zodat deze weging overbodig is.

Gezien het grote aantal punten van overeenstemming van de kenmerken van de steekproef en van de populatie, mag worden verwacht, dat de hierop gebaseerde ijking inderdaad bij zeer goede benadering landelijk representatief zal zijn.

6.3. *De normering van de test*

De normeringsprocedure verloopt, nadat langs mechanische weg de frequentieverdeling is verkregen, in principe eenvoudig. De ervaringen met de constructie van de niveauschalen hadden reeds aanleiding gegeven, de data betreffende jongens en meisjes afzonderlijk te behandelen; inspectie van de frequentieverdelingen bevestigde, dat dit terecht was geschied. Het was dus noodzakelijk, afzonderlijke normtabelletten voor jongens en meisjes samen te stellen.

Het ligt voor de hand, dat men bij een test met verscheidene schalen alle ruwe scores zal transformeren op één standaardverdeling. Hiervoor kiest men bij belangstellingstests veelal de percentielverdeling, wij hebben eerder reeds uiteengezet, dat deze rechthoekige verdeling o.i. praktische voordelen bezit. Ook IRLE heeft ogenschijnlijk een percentielverdeling gebruikt. Uit gegevens over de uitvoering van zijn ijkingsprocedure bleek ons echter, dat hij in werkelijkheid een transformatie op een normale verdeling heeft toegepast, waarbij alleen de schaalpunten niet op de gebruikelijke wijze in z- of T-scores zijn aangegeven, maar in procentwaarden. Teneinde de internationale vergelijkbaarheid van testgegevens te behouden, hebben wij deze procedure eveneens gevolgd. Achteraf kan worden gezegd, dat aan de keuze van een normale verdeling ook een belangrijk voordeel is verbonden, omdat het nu mogelijk is op veel eenvoudiger wijze het optreden van significante verschillen in de voorkeur te onderzoeken. Hierop zal in de volgende paragraaf nog worden ingegaan.

De normtabelletten voor jongens respectievelijk meisjes zijn in de afbeeldingen 6.1 en 6.2 weergegeven. Zij zijn verkregen door de ijkingsgegevens per schaal grafisch op de gekozen verdeling uit te zetten en vervolgens een vloeiende curve door de ruwe waarden te tekenen.

In de tabelletten is geometrisch een normale verdeling aangebracht, waarbij de breedte van de standaarddeviatie op 3.5 cm is gesteld. Aan de rand van de tabelletten zijn de percentielwaarden aangegeven, die nu vanzelfsprekend geen lineaire verdeling meer vertonen.

Teneinde de tabelletten op boekformaat te kunnen reproduceren, zijn niet alle mogelijke ruwe scores opgenomen. Hun plaats in het profiel kan door interpolatie worden bepaald. Indien de afstand tussen twee schaalpunten klein is en er slechts enkele ruwe scores in het betreffende gebied ontbreken, kunnen deze het beste worden ingetekend, door ze samen te nemen met de naastbijliggende waarde, die wel in de tabelletten voorkomt. Indien b.v.

BIT													Normentabel voor jongens												
%	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M	%													
99	56	50	57	40	67	33	48	57	54	30	45	99													
	55	47	55		66	32				29															
	54	44	53	38	65	31	44	53	50		43														
98		41				30		50	47	28	42	98													
+2σ	53	38	50	36	63	29	40	47				+2σ													
97	52	36		34	61	28	37	44	42	27	40	97													
		34	47		59	27																			
	50			32	56	26	34	41	36	26	38														
95		32	44			25		38	32			95													
	48		41	30	53		30	35	29	25	35														
93		30		28	50	23	26	32	27	24	33	93													
	46		38		47				25		31														
90	44	28		26		22	23	29	23	23		90													
		27	35		44				23		29														
	42	26		24	41	21	20	27	21	22															
85	+1σ	25	32		41			25	19	22	27	85													
	40	24			38	20	18	23		21	25	+1σ													
80		23	29	22			16	21	17	20		80													
	38				35	19			16		23														
75		22	27		33	15	15	19		19		75													
	36	21		20		14	14	18	15																
70		20	25		31	18	17	17	14		21	70													
	34			19		13		16	13	18															
	32	19	23		29	17		15	12		19														
60		18	21		27	16	11	14	11	17		60													
	27	16			25	15	10	13	10	16	17														
	25	15	19	16		14	9	12	9																
50		14	17	15	23	13	8	11	8	15	15	50													
	23				21		7	10	7																
	21	13	15	14		12	6	9	6	14	13	40													
40		12	13	13	19		7	8	5																
	19	11			17					13															
	17		12	12		11	6	7																	
30		10	11		15		5	6	4			30													
	15						4																		
25		9	10	11	13	10		6				25													
	11	8	9	10		9	4	5	3		9														
20		7	8		11	8			2	10	8	20													
-1σ	9	6	7	9	9		3	4				-1σ													
15		5	6	8	8	7	3	3	1	8															
	7		5	6	7					7															
	6				6																				
10		4	4	4	5		6				6	10													
	5							2																	
7		3	3								5	7													
	4			6	4																				
5								1	1			5													
	3			5	3	5					4														
3		2	2									3													
-2σ					2	4						-2σ													
2				4								2													
	1						3		0																
1		1	1	3	1	2	0			2		1													

Afbeelding 6.1

BIT		Normentabel voor meisjes										
%	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M	%
99	16	52	20	35	52	27	33	50	67	27	36	99
		51	19	34	50	25	28	48		26	33	
98	15	50	18	33	48	23	24	46	66	25	32	98
+2σ			17	32	46	22	22	44		24	31	+2σ
97	14	48	16	31	44	21	21	42	65	23	30	97
	13	46	15	30	42	20	19	40		22	29	
95	12	44	14	29	39	19	18	38	63	21	28	95
	11	42	13	28	36	18	17	36	61	20	27	
93	10	40	12	27	34	17	16	34	59	19	26	93
	9	39	11	26	32	16	15	32	57	18	25	
90	8	37	10	25	30	15	14	30	55	17	24	90
	7	35	9	24	28	14	13	28	53	16	23	
+1σ			8	23	26	13	12	26	51	15	22	+1σ
85	6	33	7	22	24	12	11	24	49	14	21	85
	5	31	6	21	22	11	10	22	47	13	20	
80	4	29	5	20	20	10	9	20	45	12	19	80
	3	27	4	19	18	9	8	19	43	11	18	
75	2	25	3	17	16	8	7	17	41	10	17	75
	1	23	2	16	14	7	6	16	39	9	16	
70		21	1	15	12	6	5	15	37	8	15	70
		19		14	10	5	4	14	35	7	14	
60		17		13	9	4	3	13	33	6	13	60
		16		12	8	3	2	12	31	5	12	
50		15		11	7	2	1	11	29	4	11	50
		14		10	6	1		10	27	3	10	
40		13		9	5			9	25	2	9	40
		12		8	4			8	23	1	8	
30		11		7	3			7	21		7	30
		10		6	2			6	19		6	
25		9		5	1			5	17		5	25
		8		4				4	15		4	
20		7		3				3	13		3	20
-1σ		6		2				2	11		2	-1σ
15		5		1				1	9		1	15
		4							8			
10		3							7			10
		2							6			
7		1							5			7
									4			
5									3			5
									2			
3									1			3
-2σ												-2σ
2												2
1												1

Afbeelding 6.2

van de ruwe scores 37 t.m. 41 er slechts twee in de tabel konden worden opgenomen, zijn hiervoor gekozen de waarden 37 en 40. De waarde 38 tekene men dan in het profiel op dezelfde plaats als 37, de waarden 39 en 41 op dezelfde plaats als 40. Dit geldt dus speciaal voor de gevallen, waarin de interpolatie, in verband met de kleine afstand tussen de schaalpunten praktische moeilijkheden zou kunnen opleveren; is de afstand groter, dan kan men ontbrekende waarden natuurlijk gemakkelijk grafisch interpoleren.

6.4. *Significantie van niveauverschillen*

Bij de psychologische interpretatie van profielen, zoals deze ontstaan door de niveauverschillen in de voorkeur voor de items van een belangstellingstest of ook door de verschillende prestaties in de subtests van een intelligentietest, komt men steeds voor de vraag te staan, of het aangetroffen verschil wellicht door toevallige variaties tot stand zou kunnen zijn gekomen. In de twintiger jaren heeft men sterk de nadruk op de betekenis van de toevalsfactoren gelegd (⁷³), ook al omdat men aanvankelijk zeer hoge eisen aan het niveau van significantie stelde. In de latere fase, toen de klinische richting in de psychologie begon te overheersen, heeft men de aanvankelijke, overdreven voorzichtigheid terzijde gelaten en is men begonnen met de duiding van profielen, waarmee men vaak waardevolle resultaten behaalde. Dit geldt speciaal voor het intelligentie-onderzoek (³⁸, ¹³⁷); bij het belangstellingsonderzoek heeft men zich over de invloed van toevallige variaties in het algemeen weinig bekommerd. Na het aanvankelijk enthousiasme voor de resultaten van de profielanalyse is ook weer een reactie gekomen van degenen, die weer oog kregen voor de foutenbronnen die hier bestaan. RAPAPORT (¹⁰⁷) komt de verdienste toe, getracht te hebben door statistische analyse een basis te vinden waarop een verantwoorde profielinterpretatie mogelijk zou zijn. Het doet in dit verband niet ter zake, dat verschillende van zijn conclusies, met name door gebreken van zijn ijkingsgroep en door het ontbreken van een kruisvalidatie, weinig houdbaar zijn gebleken. Systematische arbeid op dit terrein is sedert nog weinig verricht, wel zijn echter in de Verenigde Staten bij enkele testseries (o.a. ¹¹, ¹⁹) aanwijzingen verstrekt, waardoor het mogelijk is, op eenvoudige wijze de significantie van niveauverschillen in testprofielen te beoordelen.

Over de theorie van de toetsing van significantieverschillen tussen profielscores heeft DAVIS (⁴²) een uiteenzetting gegeven. Voor twee tests, die gestandaardiseerd zijn op dezelfde normale verdeling geldt, dat de standaarddeviatie van het verschil in score, onder de nulhypothese, gelijk is aan

$$s_v = s \sqrt{(2 - r_a - r_b)}$$

Hierin is s_v de standaarddeviatie van het verschil, s de standaarddeviatie van de normale verdeling en r_a en r_b de betrouwbaarheid van test a

respectievelijk test b. De afleiding van deze formule is gebaseerd op een analyse van de variantie van het verschil, in termen van de bijdrage van de ware score en de scorefout. Na toepassing van enkele bekende stellingen uit de test-theorie volgt dan de hierboven weergegeven formule.

Hoewel deze formule over het gehele gebied van de verdeling geldig is, betekent dit niet, dat ook over het gehele gebied de standaarddeviatie van het verschil dezelfde is. De betrouwbaarheid van de testcores behoeft niet overal dezelfde te zijn, b.v. omdat bij extreme scores de invloed van toevallige variaties anders is dan in het gemiddelde gebied. LORD ⁽⁹⁴⁾ heeft er op gewezen, dat een zuivere schatting van de zogenaamde standaard meetfout wordt gegeven door de grootheid

$$s_{MF} = \sqrt{\frac{t_a(n-t_a)}{n-1}}$$

Hierin is t_a/n het deel van het totale aantal (n) items, waarop een +reactie werd gegeven. Met behulp van deze formule kan men voor alle deelgebieden van de scoreverdeling de standaard meetfout berekenen. Beschikt men over een parallel-test, dan kan deze grootheid ook experimenteel worden gevonden.

THORNDIKE en LORGE, die dit bij de constructie van hun intelligentietest hebben gedaan ⁽¹²⁹⁾, vonden daarbij echter waarden, die over grote gebieden van de scoreverdeling slechts weinig varieerden. Alleen bij zeer hoge en zeer lage scores werden aanmerkelijk afwijkende waarden geconstateerd. Hier speelt echter de afwijkende grootte van de standaard meetfout o.i. desondanks bij de interpretatie geen rol van betekenis, omdat het feit van het optreden van een extreme score van veel meer gewicht is, dan de exacte grootte van deze score. Wij zullen daarom in het vervolg handelen, alsof de standaard meetfout over het gehele gebied dezelfde is.

DAVIS adviseert, om bij profielanalyse de critische ratio zo te kiezen, dat een significantieniveau van 15% wordt bereikt. Niveauverschillen, die groter zijn dan 1.44 maal de standaarddeviatie van het verschil, kunnen dan als significant worden geaccepteerd. Voor ons geval is het dus hier van voordeel, dat bij de ijking een normale verdeling is aangehouden, omdat dan de formule van DAVIS op alle niveau's toegepast kan worden. We behoeven dus de berekening maar eenmaal uit te voeren.

Aangezien de betrouwbaarheid van de schalen verschillend is, zou de berekening van de standaarddeviatie van het verschil voor alle combinaties van twee schalen en bovendien voor de beide sexen afzonderlijk moeten geschieden. Er zouden dan 110 waarden voor de standaarddeviatie van het verschil en dus ook voor de bijbehorende significantiegrenzen worden gevonden. Het rekenwerk aan deze bepaling verbonden is betrekkelijk gering, maar voor de practijk zou het resultaat niet te hanteren zijn. Daarom is ter vereenvoudiging voor de beide sexen een gemiddelde waarde van de betrouwbaarheidscoëfficiënt berekend en met behulp hiervan de signifi-

cantiemarge. Voor de jongens bedraagt de gemiddelde betrouwbaarheid volgens KUDER-RICHARDSON .905, voor de meisjes .815. De standaarddeviatie van de normale verdeling heeft een breedte van 3.5 cm. Uit de formule van DAVIS volgt dan, dat bij de jongens een niveauverschil van meer dan 22 mm en bij de meisjes een niveauverschil van meer dan 30 mm significant kan worden geacht. Als eenvoudig hulpmiddel zou men twee linealen van de genoemde breedten kunnen vervaardigen, waardoor men onmiddellijk visueel kan nagaan, of significantie aanwezig is.

6.5. Toevalsscores

Een bijzonder geval treedt op, als een onderzochte persoon het formulier min of meer lukraak invult. Dit geval kan zich in de eerste plaats voordoen bij zwakbegaafde of weinig-geïnteresseerde individuen, bovendien zou het ook in geval van sabotage van de test kunnen voorkomen. De vraag is, of een dergelijke toevallige invulling te herkennen is.

We beschouwen daartoe het meest extreme geval, waarin de beantwoording van de test volkomen aselekt geschiedt. In dat geval kunnen de kenmerken van de schaalscoreverdeling theoretisch worden berekend, op grond van de volgende overweging.

Op ieder kruispunt staan vier items ter keuze, die bij aselechte beslissing allen dezelfde kans hebben te worden gekozen. Voor iedere item bedraagt dus op ieder kruispunt de kansverwachting, dat deze zal worden gekozen, $\frac{1}{4}$. De beslissing, die op een gegeven kruispunt wordt genomen, heeft geen enkele invloed op de beslissingen, die op andere kruispunten worden genomen. D.w.z. dat voor de schaalscores geldt, dat zij zijn samengesteld uit de som van een aantal onderling onafhankelijke kruispuntscores. Hiervoor geldt een binomiale verdeling. Aangezien iedere schaal in totaal op 72 kruispunten ter keuze komt, zijn volgens bekende formules:

$$\mu = np = 18 \quad \sigma = \sqrt{npq} = 3.67 \quad (n = 72; p = \frac{1}{4}; q = \frac{3}{4})$$

Van de binomiale verdeling is bekend ⁽⁶²⁾, dat deze bij een groot aantal elementen de normale benadert. Dit is bij 72 kruispuntscores reeds in zeer sterke mate het geval.

Wij hebben dit nog experimenteel getoetst, d.m.v. aselechte beantwoording van een aantal testformulieren. Dit geschiedde met behulp van een tabel van „random numbers”; de cijfers 0 en 1 werden toegevoegd aan de item links boven een kruispunt, 2 en 3 aan de item rechts boven, 4 en 5 aan de item links beneden en 6 en 7 aan de item rechts beneden. De cijfers 8 en 9 werden niet gebruikt. Op ieder kruispunt kon nu met behulp van genoemde tabel worden vastgesteld, welke item zou worden gekozen. In totaal werden 25 formulieren op deze wijze ingevuld en vervolgens gescoord. Aangezien de negen schalen uiteraard bij aselechte invulling als onderling verwisselbaar kunnen worden behandeld, werd op deze wijze

de beschikking verkregen over 225 toevalsscores op de schalen. De gemiddelde score was — vanzelfsprekend — 18, terwijl de standaarddeviatie 3.55 bleek te zijn. De gevonden verdeling bleek niet significant te verschillen van een normale verdeling, met de bovengenoemde theoretisch berekende kenmerken.

We kunnen nu deze toevalsscores in het testprofiel aangeven. Dit is gebeurd door een klein gebied rondom de theoretisch verwachte toevalscore te schaduwen; door het gebied enige breedte te geven, kan worden voorkomen, dat verwarring met een ingetekende profiellijn optreedt.

Uit de ligging van de geschaduwde gebieden blijkt, dat iemand, die de test geheel aselekt zou beantwoorden, geen doorsnee-profiel krijgt. In tegendeel zou een dergelijke persoon ogenschijnlijk een zeer karakteristiek interessenprofiel vertonen. Hoe meer nu een gevonden profiel het toevalsbeeld benadert, des te groter is de kans, dat de betrokken persoon lukraak heeft geantwoord. Weliswaar is het denkbaar, dat iemand bij serieuze invulling ditzelfde beeld zou vertonen; men zal dus steeds de veronderstelling, dat men met een lukrake beantwoording te maken heeft moeten verifiëren, b.v. door middel van interview-gegevens.

Ter toetsing van de veronderstelling, dat bij de invulling van een gegeven testformulier toevallige invloeden een zeer sterke rol hebben gespeeld, kan men een procedure volgen, die gebaseerd is op de teken-toets ⁽⁷¹⁾. Grondgedachte is, dat de kans, dat de score op een gegeven schaal boven de mediaan ligt, even groot is als de kans, dat deze beneden de mediaan ligt. De kans, dat het geschaduwde gebied aan dezelfde zijde van de mediaan ligt als de gegeven schaalscore is dus $\frac{1}{2}$. Indien nu tenminste tien van de elf schaalscores van een gegeven persoon aan dezelfde zijde van de mediaan liggen als de geschaduwde gebieden, moet serieus rekening worden gehouden met de mogelijkheid van lukrake beantwoording. Ook indien dit voor negen schaalscores het geval is, verdient het aanbeveling, deze mogelijkheid niet geheel buiten beschouwing te laten. Waar het geschaduwde gebied ongeveer samenvalt met de mediaan zal een gemiddelde score moeten worden opgevat als een aanwijzing in de bovenbedoelde zin.

Vanzelfsprekend moet men, behalve met de richting van de afwijking van de mediaan, ook rekening houden met de afstand van de schaalscore tot het geschaduwde gebied. Is deze afstand, in score-eenheden uitgedrukt, zeer groot, dan spreekt dit duidelijk tegen de veronderstelling, dat de onderzochte lukraak zou hebben geantwoord.

HOOFDSTUK VII

VAN INTERESSETEST TOT INTERESSEKARAKTERISTIEK

7.1. *Het validatieprobleem*

Het eerste deel van ons onderzoek heeft betrekking gehad op de bewering, toetsing en ijking van de B I T. In het thans te bespreken tweede deel van de studie is deze test gebruikt voor een onderzoek naar de karakteristieke reacties van een aantal groepen, die op het overgangspunt van het lager naar het voortgezet onderwijs kunnen worden onderscheiden.

Door dit gebruik van de test worden we geconfronteerd met het probleem van de validiteit van onze methode. Op deze problematiek zal eerst moeten worden ingegaan.

Tot voor enkele jaren lag het validiteitsprobleem ogenschijnlijk betrekkelijk eenvoudig. In de literatuur werden twee begrippen gehanteerd. Het ene, validiteit zonder meer genoemd, had betrekking op de empirisch en met statistische middelen bepaalde overeenstemming van de testgegevens met een uitwendig criterium. Het andere, „face validity”, hetgeen wij willen vertalen met uiterlijke validiteit, had betrekking op een soortgelijke overeenstemming, die echter subjectief, op grond van het intuïtieve, klinische oordeel van de psycholoog, aan de test werd toegekend. KOUWER (77) heeft er op gewezen, dat de eerstgenoemde betekenis van het validiteitsbegrip in ons land tot dusver weinig gehanteerd is.

De laatste jaren hebben echter de opvattingen over het validiteitsbegrip een aanzienlijke evolutie doorgemaakt, waardoor het beeld veel gecompliceerder is geworden. In dit verband moet in de eerste plaats het werk van het „A P A Committee on psychological tests” (18) genoemd worden. Deze heeft een onderscheiding gemaakt in „content”-, „construct”-, „concurrent”- en „predictive validity”. Aan deze begrippen heeft VAN DER GIESSEN (50) onlangs reeds een uitvoerige beschouwing gewijd, zodat een toelichting hier achterwege kan blijven. Wij wijzen er alleen op, dat de onderscheidingen van de Amerikaanse studietoelichting gewijd, zodat een toelichting hier achterwege kan blijven. Wij wijzen er alleen op, dat de onderscheidingen van de Amerikaanse studietoelichting gewijd, zodat een toelichting hier achterwege kan blijven. Wij wijzen er alleen op, dat de onderscheidingen van de Amerikaanse studietoelichting gewijd, zodat een toelichting hier achterwege kan blijven.

Het aantal validiteitsbegrippen is sedert nog weer toegenomen. In de Amerikaanse literatuur komt ook nog het begrip „congruent validity” (129) voor, waarmee de mate van overeenstemming van de testgegevens, met gegevens afkomstig van een soortgelijke, reeds gevalideerde test, wordt bedoeld. Terwijl dit een waardevolle aanvulling van de begrippenschat is,

werkt daarentegen de poging van anderen, om de voorgestelde namen door andere te vervangen, alleen verwarrend. Zo heeft NOLL^(103a, geciteerd bij⁸⁷) getracht, de termen „curricular”-, „logical”- en „empirical validity” te introduceren, die respectievelijk met het eerste, tweede en gecombineerde derde en vierde validiteitsbegrip van de Amerikaanse studiec commissie corresponderen. Enige winst kan van een dergelijke naamsverandering niet worden verwacht. Wel verdient het o.i. aanbeveling, te trachten Nederlandse termen te vinden ter aanduiding van de Amerikaanse equivalenten. In deze studie zullen wij in het vervolg spreken van respectievelijk inhoudelijke, begrips-, synchrone, predictieve en congruente validiteit, ter aanduiding van de vijf in dezelfde volgorde genoemde Amerikaanse begrippen.

Bij al deze vormen van validiteit hebben wij in feite te maken met de relevantie van bepaalde aspecten van de test. DE GROOT⁽⁵⁵⁾ wijst er op, dat de verschillende vormen in wezen tot twee categorieën van validiteit behoren: de relevantie van de test in psychologisch opzicht en de relevantie in praktisch opzicht. Tot de eerstgenoemde groep behoren de inhoudelijke en de begripsvaliditeit. Tot de tweede categorie behoren de synchrone en de predictieve validiteit. De onderscheiding tussen deze beide laatste begrippen heeft uitsluitend praktische betekenis. In wezen is er geen verschil tussen het geval, dat het criterium, waarover op grond van de testuitkomsten een uitspraak wordt gedaan, in het verleden, het heden of de toekomst ligt. Maar voor degene, die validiteitsonderzoek verricht, is de onderscheiding desondanks nuttig, zowel wegens de praktische consequenties voor het onderzoek, als omdat de beoordeling van de validiteit wellicht anders zal zijn, wanneer tussen de beide reeksen te vergelijken gegevens een zekere tijdsafstand ligt. Op grond van deze overweging verdient het o.i. aanbeveling, de onderscheiding toch te handhaven. De term congruente validiteit kan, analoog aan de uiterlijke en de „algemene” validiteit, in verband met al de vier genoemde vormen worden gebruikt, zij kan dus ook zowel met betrekking tot de psychologische, als tot de praktische relevantie worden toegepast.

Welke betekenis hebben deze validiteitsbegrippen in verband met onze studie? De inhoudelijke validiteit behoeft in statistische zin bij een belangstellingstest geen rol te spelen. Het is voor een goede belangstellingstest niet essentieel, dat hij naar inhoud een representatieve steekproef uit de te toetsen interessegebieden bevat. Houdt men echter ook rekening met de uiterlijke validiteit, dan wordt de situatie in ons geval wel iets anders. Hoewel ook thans de aanwezigheid van inhoudelijke validiteit nog steeds geen gebiedende eis is, kan men toch zeggen, dat de B I T bestaat uit items, die ogenschijnlijk wel in de gekozen gebieden thuishoren.

De begripsvaliditeit speelt bij onze test een zeer belangrijke rol, zodanig, dat hieraan in de volgende paragraaf een uitvoeriger beschouwing zal worden gewijd.

Aangezien wij de test willen gaan gebruiken in het kader van de selectie aan het einde van de lagere school, speelt het praktische aspect van de validiteit vanzelfsprekend een zeer grote rol. In de meeste gevallen hebben wij ons beperkt tot een vergelijking van de testuitkomsten met een gelijktijdig gegeven criterium, zodat dus voornamelijk de synchrone validiteit in het geding is. In één geval is echter een aantal onderzochte kinderen gedurende één jaar gevolgd, teneinde ook hun schoolresultaten in onze studie te kunnen betrekken. Voor zover daarbij gebleken is, dat de testuitkomsten relevant zijn voor de voorspelling van het schoolsucces, kan dus van predictieve validiteit worden gesproken.

In ons onderzoek komt de congruente validiteit nauwelijks aan de orde, omdat soortgelijke tests in ons land nog niet beschikbaar zijn. Alleen in één geval, bij de sexe-karakteristiek, zal een poging worden gedaan, met behulp van gegevens ontleend aan een buitenlandse test, tot een congruente validatie te komen.

7.2. Operationele definitie

In paragraaf 1.1 hebben wij het begrip interesse gedefinieerd als „het aspect van de motivatie, dat bestaat uit de bewust-beleefbare gerichtheid op bepaalde categorieën emotioneel stimulerende en tot actie uitlokkende situaties”. Wij hebben er toen reeds op gewezen, dat een dergelijke definitie zich in deze vorm slecht leent voor een empirisch onderzoek. Zij zal eerst operationeel moeten worden gemaakt. Bij voorkeur zal men deze operationele definitie zonder meer uit de oorspronkelijke willen afleiden, gebruik makende van de aspecten van de begripsomschrijving, die verwijzen naar objectief waarneembare entiteiten. In het geval van het interessebegrip zou dit kunnen geschieden door in de operationele definitie gebruik te maken van de op belangstellingsstimuli gerichte actie, d.w.z. dus van de manifeste interessen. We hebben echter in het eerste hoofdstuk reeds gezien, dat de ervaringen met het gebruik van deze interessen voor psychodiagnostische doeleinden, i.v.m. school- of beroepskeuze, niet zeer gunstig zijn geweest. Dit is aanleiding geweest, gebruik te gaan maken van „inventoried interests”.

Hier doet zich echter een moeilijkheid voor, want het belangstellingsbegrip, waarmee in vragenlijsten wordt gewerkt, kan niet zonder meer uit de boven gegeven definitie worden afgeleid. Het feit b.v., dat iemand een bepaalde item uit onze belangstellingstest herhaaldelijk kiest, betekent nog niet, dat hij in de realiteit ook deze activiteit zal gaan verrichten, wanneer hij in de gelegenheid zou komen, het te doen. Er ligt tussen het op papier uitspreken van een voorkeur en het werkelijk tot actie komen, nog een zeer grote afstand. We mogen de belangstelling, zoals deze met behulp van de test wordt gemeten, dus niet zonder meer als een operationele vorm van het boven gedefinieerde interessebegrip beschouwen. De vraag, die

ons dus moet bezighouden is, of het „construct” interesse, zoals boven gedefinieerd, inderdaad op valide wijze door onze test wordt gemeten. De beantwoording van deze vraag is niet eenvoudig. Men kan zich op het standpunt stellen, dat in de homogeniteitsanalyse (vergelijk paragraaf 5.2) een poging is gedaan om de consistentie van de schalen aan te tonen en dat daarmee ook een bijdrage is geleverd tot het bewijs van de begripsvaliditeit. Als de innerlijke consistentie ontbreekt, zal het interessebegrip, dat bij de constructie van de schaal is gehanteerd, zeker gebreken vertonen. Homogeniteit van de items met de schaal is dus een noodzakelijke voorwaarde voor begripsvaliditeit. Maar daarnaast zullen we ook moeten aantonen, dat de testcores een voldoende correlatie vertonen met langs andere weg verkregen belangstellingskarakteristieken. Daartoe zou men b.v. de congruente validiteit met andere belangstellingstests, waarvan de begripsvaliditeit reeds voldoende vaststaat, willen bepalen. We hebben echter gezien, dat deze weg in het algemeen voor ons is afgesloten, omdat dergelijke tests nog niet ter beschikking staan. Hetzelfde geldt voor een andere validatiemogelijkheid, namelijk de testuitkomsten betreffende de beoefenaren van een aantal beroepen te vergelijken met de reeds bekende belangstellingskarakteristiek van deze beroepen. Ook IRLE is bij zijn onderzoek over de B I T op deze moeilijkheid gestuit, toen hij moest constateren, dat dergelijke karakteristieken nog niet bestaan. Dit is voor hem zelfs aanleiding geweest, op het validiteitsprobleem niet dieper in te gaan.

Analyseert men het bovengenoemde probleem verder, dan blijkt, dat wij in een vicieuze cirkel dreigen te geraken. De test kan alleen worden gevalideerd aan de hand van bekende belangstellingskarakteristieken of van reeds gevalideerde tests. Deze kunnen echter op hun beurt alleen worden verkregen met behulp van een test, waarmee valide uitspraken kunnen worden gedaan. Want het menselijk oordeel schiet in deze situatie, door zijn subjectiviteit, zonder de hulp van dergelijke instrumenten tekort.

Er moet dus een middel worden gezocht, om deze kringloop te doorbreken. Men zou kunnen overwegen, de moeilijkheid op te lossen, door af te zien van een algemeen-psychologische begripsbepaling en het belangstellingsbegrip eenvoudig operationeel te definiëren als datgene, wat met behulp van de schalen van de test wordt gemeten. Een dergelijke arbitraire oplossing is echter niet bevredigend. Het geringste telt voor ons nog het bezwaar, dat er bij vele Nederlandse psychologen, om van de ruimere kring van belangstellenden, b.v. in de onderwijswereld nog maar te zwijgen, toch reeds weinig sympathie voor het gebruik van operationele definities bestaat. Maar het essentiële bezwaar tegen deze werkwijze is, dat we hier te maken krijgen met een operationele definitie achteraf.

Bij de constructie van de test is immers uitgegaan van een belangstellingsbegrip, dat ongetwijfeld via introspectie en observatie verband houdt met het algemeen-psychologische belangstellingsbegrip. Gaan we nu achteraf, in onze operationele definitie, dit verband loslaten, dan is het gevaar

groot, dat de testgebruiker toch de uitkomsten gaat interpreteren in termen van het inzichtelijk verkregen interessebegrip.

Ditzelfde probleem doet zich ook voor bij andere begrippen, die operationeel worden gedefinieerd. Naar onze mening zijn velen, die een voorkeur hebben voor dergelijke definities — die voor de practijk van het experimenteel onderzoek van essentieel belang zijn — geneigd, de gevaren die in het gebruik van een operationele definitie kunnen schuilen, te negeren. Er schuilt een zekere mate van intuïtieve juistheid in de critiek van hen, die de operationele definitie een te gemakkelijke oplossing van de moeilijkheden vinden. We kunnen in dit verband verwijzen naar de ervaringen met het intelligentieonderzoek. Ook daar is men gestoten op het verschijnsel, dat „het criterium voor de objectiviteit en de methodische juistheid van (de) test wordt gevonden met behulp van iets, dat men vanwege de er aan verbonden onvermijdelijke subjectiviteit, verwerpt, namelijk de schoolprestaties van het kind en de beoordeling daarvan door de school” (115, pag. 54). Inderdaad schijnt het, alsof de test, die het subjectieve oordeel wil corrigeren, van datzelfde oordeel afhankelijk blijft. Dit bezwaar blijft gelden, ook als we achteraf overgaan op een operationele definitie.

Toch is er een uitweg, die onder anderen is gewezen door WECHSLER (137), DE GROOT (56) en CRONBACH en MEEHL (40). Laatstgenoemden spreken van het bestaan van een „bootstraps-effect”, hetgeen wij hier met „*Münchhausen effect*” zouden willen vertalen. De intelligentietest b.v., die we aanvankelijk alleen durfden gebruiken omdat er een redelijke mate van overeenstemming tussen de testuitkomsten en het subjectieve oordeel van onderwijzers over de begaafdheid van hun leerlingen bestond, krijgt later een zelfstandige positie. Hij kan nu dienen als toetssteen voor diezelfde oordelen. Schijnbaar heeft de test zich, als *von Münchhausen* aan de eigen haren uit het moeras getrokken.

In werkelijkheid heeft echter een volkomen legitiem verklaarbare ontwikkeling plaats gevonden, die neerkomt op het zuiveren van de onderzoeksmethode. Al dan niet uitgesproken uitgangspunt van deze procedure is geweest, dat de gemiddelde mens beschikt over een zekere mate van intuïtieve mensenkennis, waardoor zijn oordeel over een bepaalde psychische kwaliteit, bij personen die hij vrij goed kent, enige begripsvaliditeit bezit. Zou de methode, die we hebben ontworpen, in het geheel geen begripsvaliditeit bezitten, dan zou dit moeten blijken uit het totaal ontbreken van een correlatie met de beoordeling door de genoemde mensenkenners. Er is echter correlatie, dus ook een aanduiding van validiteit. Heeft men aldus eenmaal een basis gevonden, dan kan men de test methodisch verbeteren, o.a. door het elimineren van items, die niet met het totaal correleren, door het vergroten van de betrouwbaarheid, een goede ijking, etc. Bovendien kan men er gelijktijdig naar streven, ook een beter theoretisch inzicht te krijgen in het begrip, dat we met behulp van de test hebben

getoetst. De test krijgt op deze wijze een beter differentiërend vermogen en mede daardoor ook een validiteit, die veel groter is, dan die van het oordeel van de gemiddelde mensenkenner. M.a.w. de test is nu de beste beschikbare maatstaf voor de psychische kwaliteit in kwestie geworden; hij kan daarom dienen als norm voor de toetsing van het oordeel dergenen, die aanvankelijk door ditzelfde oordeel een rechtvaardiging voor de juistheid der testconstructie hebben geleverd.

Omdat de test nu de beste beschikbare maatstaf voor de betreffende kwaliteit is geworden, mogen we vervolgens overgaan tot een operationele definitie. De vicieuze cirkel wordt op deze wijze dus doorbroken door een combinatie van twee factoren: enerzijds het beroep op het bestaan van een, zij het eventueel ook bescheiden mate van intuïtieve mensenkennis, anderzijds door het gebruik dat wordt gemaakt van de methoden tot testverbetering, met o.a. als gevolg de uitschakeling van subjectieve invloeden. Op deze wijze gezien kan de operationele definitie indirect toch in verband worden gebracht met de oorspronkelijke begripsbepaling.

Het *Münchhausen-effect*, dat bij de intelligentietests met veel succes is gebruikt, kan ook in de methodiek van het interessesonderzoek van belang zijn. Ook bij dit laatste wordt begonnen met een beroep te doen op de intuïtieve mensenkennis, ditmaal door intuïtief de items te kiezen, die representatief worden geacht voor bepaalde schalen. In een nog vroeger ontwikkelingsstadium zijn ook deze schalen intuïtief gekozen, de juistheid van de keuze is echter bevestigd door de grote mate van overeenstemming in de uitkomsten van pogingen, die op zeer verschillende wijze waren opgezet. Via een proces van testzuivering komt men vervolgens geleidelijk tot een test, die onafhankelijk van het oordeel van de incidentele beoordeelaar als maatstaf voor de interesses kan dienen. We kunnen dan een dergelijke test ook gebruiken als hulpmiddel om de belangstellingskarakteristiek van een bepaalde gegevenheid, b.v. die van de gemiddelde beoefenaar van een beroep of van de gemiddelde leerling van een opleiding te leren kennen. Het is voldoende, dat we daartoe aantonen, dat er een zekere mate van initiële overeenstemming tussen de testuitkomsten en een intuïtieve beoordeling aanwezig is. We beginnen dus een hecht samenhangend geheel te construeren, om dit vervolgens in de realiteit te verankeren. Gelukt dit, dan mogen we de testuitkomsten ook gebruiken in een operationele definitie, het gevaar van een misplaatste generalisatie van de testgegevens door de gebruiker, is nu sterk verminderd.

Beschikken we eenmaal over een belangstellingstest, die behoorlijk gevalideerd is, dan levert de validatie van volgende tests minder problemen op, omdat men dan de eerste als criterium kan gebruiken. Nu een dergelijke test in Nederland nog niet bestaat, zullen we echter empirisch moeten aantonen, dat in een voldoende aantal situaties inderdaad overeenstemming tussen de testuitkomsten en een intuïtief gegeven of intuïtief aanvaardbare karakteristiek aanwezig is.

7.3. De belangstellingskarakteristiek

In de praktijk is het lastig, de hiervoor bedoelde begripsvalidatie scherp gescheiden te houden van de toepassing, die gebaseerd is op de veronderstelling, dat de validatieprocedure reeds tot een succesvol einde is gevoerd. Dikwijls zal eenzelfde situatie de mogelijkheid bieden, aan beide opgaven aandacht te schenken. Tendele zal daardoor het gebruik van de test, voor het bepalen van de belangstellingskarakteristiek van bepaalde groepen, voorafgaan aan de validatie.

In het algemeen gesproken zal de mogelijkheid tot begripsvalidatie vooral dan aanwezig zijn, indien wij bij bepaalde groepen mensen handelwijzen kunnen waarnemen, die in overeenstemming zijn met de verwachting, die op grond van de testuitkomsten kan worden opgesteld. Indien b.v. personen, die in de test een duidelijke voorkeur voor items uit de technisch-manuele sector hebben getoond, een metaaltechnische vakopleiding kiezen of volgen, beantwoordt dit aan de inzichtelijk verkregen verwachting. Soortgelijke verwachtingen kan men ook ten aanzien van andere opleidingsmogelijkheden opstellen. Indien zij voldoende vaak worden bevestigd, helpen zij onze operationele definitie te verankeren, zowel in de uitwendige realiteit als in de algemeen-psychologische begrippenwereld.

In verband met het voorgaande moeten wij nog een scherp onderscheid maken tussen het gebruik van dezelfde situatie voor begrips- en voor synchrone validatie. Wanneer we de genoemde bevindingen ten aanzien van aspirant-metaalbewerkers doen, dan levert dit een bevestiging van onze intuïtieve verwachting op grond van het interessebegrip, maar het is tegelijkertijd een constatering van de praktische relevantie van de test voor de differentiatie na de lagere school. Praktische relevantie kan echter uiteraard ook aanwezig zijn, zonder dat er van begripsvalidatie sprake is, b.v. wanneer we constateren, dat meisjes, die v.h.m.o. kiezen, gemiddeld een zeer geringe sympathie voor commerciële activiteiten tonen. Er is geen theorie, op grond waarvan men dit kenmerk bij voorbaat kan verwachten, nochtans blijkt het als differentiërend criterium bruikbaar, het bezit een goede synchrone en zelfs ook predictieve validiteit.

Volgens DE GROOT⁽⁵⁵⁾ kan de begripsvaliditeit niet in strikte zin worden bewezen en evenmin gemeten. De begripsvaliditeit berust op een beoordeling van het totale „nomologische netwerk”, gerepresenteerd door de uitkomsten van onderzoeken, waarin het begrip of het instrument betrokken is. De begripsvaliditeit neemt dus toe met de uitvoering van validatie- en evaluatiestudies, indien deze althans een bevestiging van de uitgangshypothesen leveren.

De verder te bespreken uitkomsten van ons onderzoek leveren o.i. inderdaad het materiaal, dat ons veroorlooft, van een toenemende begripsvaliditeit te spreken. Daarom zal in het algemeen worden gehandeld, alsof de begripsvaliditeit in voldoende mate bestaat. Dit is met name van belang

bij het opstellen van de belangstellingskarakteristiek van verschillende groepen, die b.v. in het voortgezet onderwijs kunnen worden onderscheiden. We zullen de rangorde van de scores van een concreet individu beschouwen als een rechtstreekse aanwijzing over de rangorde van de sterkte der bij hem aanwezige interessen.

Analoog beschouwen we ook de rangorde van de mediale of gemiddelde schaalscores bij een groep, als een aanwijzing over de rangorde der gemiddeld aanwezige sterkten der interessen. In deze laatste werkwijze, die het ons mogelijk maakt, karakteristieke interesseprofielen voor bepaalde groepen samen te stellen, ligt nog een bijzondere veronderstelling opgesloten, namelijk, dat het geoorloofd is schaalscores van verschillende individuen cumulatief te behandelen. Ongetwijfeld zal men niet mogen veronderstellen, dat bij twee individuen, die op dezelfde schaal dezelfde ruwe score behalen, de motiverende krachten in de bedoelde interesserichting ook een gelijke sterkte hebben. Wel echter zal men, als de betreffende score b.v. aanzienlijk boven het gemiddelde ligt, mogen aannemen, dat deze interesserichting in de belangstellingshiërarchie van de beide personen een voorname rol speelt. Dit is het kenmerk, dat we bij de cumulatieve verwerking van de scores van een aantal individuen, in een belangstellingskarakteristiek, gaan hanteren.

HOOFDSTUK VIII

SEXEVERSCHILLEN IN DE BELANGSTELLING

8.1. *Vraagstelling*

Bij het onderzoek van de karakteristiek van verschillende groepen in het voortgezet onderwijs is ons gebleken, dat jongens en meisjes, die hetzelfde schooltype bezoeken, dikwijls aanzienlijke verschillen in belangstellingskarakteristiek vertonen. Ook bij een bestudering van de ijkingsstabellen vinden we grote sexeverschillen. Daarom is het gewenst aan deze verschillen aandacht te schenken, voor wij op de kenmerken van andere groepen ingaan.

Het bestaan van bepaalde sexeverschillen in de belangstelling is algemeen bekend. De voorkeur van mannen voor b.v. techniek en natuurwetenschap, zal door vakman en leek evenzeer algemeen worden erkend, als de vrouwelijke voorkeur voor het artistieke, literaire en sociale. Men kan dus zelfs als eis aan een goede belangstellingstest stellen, dat dergelijke kenmerken bij onderzoek inderdaad zullen worden gevonden. De meeste belangstellingstests houden hiermee ook bij voorbaat rekening, door verschillende normenschalen voor mannen en vrouwen te geven. Het is in feite zelfs zeer opmerkelijk, dat IRLE bij de constructie van zijn testnormen hieraan is voorbij gegaan. In zijn ijkingsmateriaal waren immers zowel mannen als vrouwen opgenomen. Doordat echter het aantal mannen sterk overwoog, mag wel worden gesteld, dat de door IRLE gegeven normen in feite op de mannelijke interessekarakteristiek betrekking hebben.

Een analyse van de sexeverschillen, die wij in ons materiaal vinden, is om twee redenen interessant. In de eerste plaats biedt een dergelijke analyse ons de gelegenheid, een beter inzicht te krijgen in de sexeverschillen in de belangstelling bij de speciale leeftijdsgroep, die door ons is onderzocht. De kennis, die wij hieromtrent bezitten, is grotendeels nog van voorwetenschappelijke aard, experimentele bevindingen zijn er weinig over gepubliceerd. Weliswaar behoeft in dit geval niet te worden verwacht, dat het experimenteel onderzoek tot conclusies zal voeren, die met de voorwetenschappelijke kennis in strijd zijn, maar wel kan deze kennis daardoor worden gepreciseerd.

In de tweede plaats bestaat ook de mogelijkheid, dat wij door deze analyse tot een congruente validatie van onze test komen. We moeten dan uitgaan van de veronderstelling, dat de richting van de sexeverschillen in

de belangstelling gedurende grote perioden van het leven constant blijft. Op basis van deze veronderstelling kunnen we de richting van de sexeverschillen in ons materiaal vergelijken met de sexekarakteristiek van een andere belangstellingstest, die voldoende verwant is om een vergelijking mogelijk te maken. In concreto kan dit geschieden, door de richting van de sexeverschillen op de *KUDER Preference Record* ⁽⁸⁸⁾ met die op onze test te vergelijken.

Van de beide genoemde problemen, zal het tweede in de volgende paragraaf worden behandeld. De meer gedetailleerde bestudering van de sexeverschillen vormt het onderwerp van het laatste deel van dit hoofdstuk.

8.2. *Congruente validatie van de sexeverschillen*

Ten behoeve van de vergelijking van de B I T met de *KUDER Preference Record* hebben wij gebruik gemaakt van de ligging van de medianen der ruwe scores van de beide sexen op ieder van deze beide tests. De gegevens over de B I T zijn ontleend aan de in hoofdstuk VI gereproduceerde normtabelletten, de gegevens over de KUDER-test aan het „profile sheet”, dat ten behoeve van de zelfbeoordeling in de handel verkrijgbaar wordt gesteld. Een nadere analyse van de normenscalen op beide tests leert, dat vrijwel steeds de ligging van de mediaan representatief is voor sexeverschillen over de gehele schaalverdeling. Indien op de mediaan een sexeverschil aanwezig is, dan wordt als regel ook op het eerste en negende deciel een gelijk gericht verschil gevonden. Wel kan echter de grootte van het verschil in de verschillende gebieden van de schaalverdeling uiteenlopen. Gebruik van de mediaan geeft dus een goede mogelijkheid, via de tekentoets een eventuele overeenstemming te constateren. Het blijkt niet mogelijk, de verwantschap nog nader te onderzoeken, door ook de grootte van de verschillen op diverse schalen in de vergelijking te betrekken, omdat verschillen in de details van de scoringswijze hier storend werken. Bovendien hebben de normen niet op dezelfde leeftijdsgroepen betrekking, zodat de uitgangshypothese niet al te specifiek gesteld moet worden.

Zoals bekend heeft de *KUDER Preference Record* tien schalen. Eén hiervan, het muzikaal interesse, moet buiten beschouwing blijven, omdat deze in de B I T niet voorkomt; omgekeerd heeft van de B I T de voedselbereidingsschaal geen tegenhanger. De toevoeging van de overige KUDER-schalen aan die van de B I T is gemakkelijk, behalve in het geval van de administratieve interessen, waaraan bij de KUDER-test zowel de „computational” als de „clerical”-schaal zouden kunnen beantwoorden. In het algemeen valt onder het laatstgenoemde begrip vooral het administratieve routinewerk, maar er is een vrij sterke overlapping met het eerstgenoemde begrip. Het probleem wordt hier gecompliceerd door het feit, dat de richting van de sexeverschillen op de KUDER-test bij deze beide schalen tegengesteld is. Daarom is besloten, het administratief interesse eveneens buiten

beschouwing te laten. De vergelijking heeft dus uiteindelijk betrekking op zeven schalen.

Tabel 8.1 geeft voor ieder der tests eerst de ligging der medianen bij de beide sexen. Vervolgens wordt met behulp van de letters j en m aangeduid, of de mediale score bij jongens respectievelijk bij meisjes hoger is. Een vergelijking van deze letteraanduidingen geeft de maatstaf voor de beoordeling van de mate van overeenstemming tussen de beide tests.

TABEL 8.1. Sexeverschillen in de B I T en de KUDER-Preference Record

	Mechanical TH	Artistic AV	Science TN	Outdoor AA	Persua- sion HA	Literary LG	Social service SO
<i>B I T</i>							
jongens	24	15	18	23	14	12	9
meisjes	2	27	3	19	9	19	37
<i>richting</i>	j	m	j	j	j	m	m
<i>KUDER</i>							
jongens	45	25	41	47	40	17	37
meisjes	22	28	28	30	39	19	48
<i>richting</i>	j	m	j	j	j	m	m

De richting van de verschillen blijkt dus in alle zeven gevallen bij beide tests dezelfde te zijn. Aangezien onze uitgangshypothese is, dat de sexekarakteristiek bij beide tests dezelfde is, mogen we eenzijdig toetsen; dit betekent, dat de gevonden overeenstemming een toevalskans van minder dan 1% heeft.

Deze bevinding levert dus in de eerste plaats een bijdrage tot de begripsvalidatie van de test. Maar zij geeft bovendien steun aan de hypothese, dat de sexekarakteristiek van de belangstelling over betrekkelijk lange perioden naar richting constant blijft. De weergegeven mediaanwaarden wijzen er op, dat reeds aan het einde van de lagere school deze karakteristiek gevormd is, zodat het ontstaan er van in een vroege levensfase moet liggen.

8.3. Sexeverschillen in de B I T

Teneinde de sexekarakteristiek van de belangstelling bij de door ons onderzochte leeftijdsgroep nader te bestuderen, dient de frequentieverdeling van de schaalscores meer in detail te worden weergegeven. De percentielverdeling van de ijkingsstabellen is, door het grote aantal onderverdelingen, moeilijk in een dergelijke analyse te hanteren. Daarom is hier de wat grovere decielverdeling gebruikt. Deze zal ook in de volgende hoofdstukken het referentiekader voor het onderzoek van de belangstellingskarakteristieken vormen, zij wordt daarom in tabel 8.2 weergegeven. De daarin opgenomen decielwaarden zijn afgeleid uit het ijkingsmateriaal, met afronding op de naastbijgelegen ruwe score.

TABEL 8.2. Frequentieverdeling der ruwe scores in decielen, voor jongens en meisjes

Schaal		D ₁	D ₂	D ₃	D ₄	D ₅	D ₆	D ₇	D ₈	D ₉
TH	j	6	10	15	20	24	29	34	39	45
	m	—	0	—	1	2	3	4	5	9
AV	j	5	8	10	12	15	17	20	23	28
	m	16	20	23	25	27	29	32	35	40
TN	j	5	9	12	15	18	21	25	29	37
	m	0	1	—	2	3	4	6	8	11
VB	j	7	10	12	14	16	17	19	22	26
	m	12	14	16	17	19	20	22	24	27
AA	j	7	12	16	20	23	27	31	36	47
	m	8	11	14	16	19	21	25	28	33
HA	j	7	9	11	13	14	16	18	20	23
	m	3	4	6	7	9	10	12	14	17
AD	j	3	4	6	8	9	11	13	17	24
	m	1	2	3	5	6	8	10	12	16
LG	j	2	5	7	9	12	14	17	22	30
	m	8	12	14	17	19	21	24	28	33
SO	j	1	3	5	7	9	11	14	18	25
	m	19	24	29	33	37	41	47	52	59
L	j	8	11	13	14	15	17	18	20	23
	m	10	12	13	14	15	17	18	19	21
M	j	6	9	11	13	16	18	21	25	31
	m	7	9	11	13	15	17	19	21	25

Bij het gebruik van de gegevens van deze tabel dienen we er ons van bewust te blijven, dat de verschillen die d.m.v. de B I T worden geconstateerd, van relatieve aard zijn. Indien b.v. uit de test blijkt, dat de items uit een bepaalde schaal door meisjes meer worden gekozen dan door jongens, mogen we hieruit niet afleiden, dat ook in absolute zin de drijvende werking van de betreffende interesserichting bij meisjes groter is. De schaalwaarden geven alleen een rangorde van de sterkten der interessen.

Dit is uiteraard een uitvloeisel van het gebruik van het principe van de gedwongen keuze.

Uit de tabel blijken onmiddellijk de grote sexeverschillen in de meeste schalen. We willen daarom in de eerste plaats de karakteristiek van de beide sexen nagaan, door de rangorde van de voorkeur voor jongens en meisjes afzonderlijk vast te stellen. Daarna zal ook een onderlinge vergelijking van de reacties der beide sexen geschieden.

Teneinde de verdeling van de scores binnen de twee sexen na te gaan, construeren we een standaardverdeling, d.w.z. de testprestatie van een hypothetische persoon, die op alle schalen een gemiddelde score behaalt.

In dit geval is het rekenkundig gemiddelde gebruikt, omdat de verdeling ook moet voldoen aan de eis, dat de som van de eerste negen schaalscores ongeveer 162 is. (Door het weglaten van enkele niet-homogene items uit de beoordeling wordt de verwachting t.a.v. de som iets lager dan 162, de juiste waarde is echter moeilijk te bepalen.) De gemiddelde waarden blijken iets hoger te liggen dan de medianen, in vele gevallen corresponderen zij ongeveer met het zesde deciel. Tabel 8.3 geeft de verdeling van de standaard voor jongens en meisjes afzonderlijk.

TABEL 8.3. Standaardverdeling (schaalgemiddelden) voor jongens en meisjes

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
jongens	25	16	20	17	25	15	12	15	12	16	18
meisjes	4	28	5	19	21	10	8	21	38	16	16

Uit de tabel blijkt, dat de verdeling van deze scores zeer aanmerkelijk afwijkt van de theoretisch berekende toevalsverdeling. Op grond van de in paragraaf 6.5 genoemde waarden zouden bij een toevalsverdeling, die betrekking heeft op ongeveer 500 personen, de gemiddelden een spreiding met een standaarddeviatie van .16 à .17 rondom de verwachte waarde van 18 moeten vertonen. De wijziging veroorzaakt door het weglaten van enkele items uit de scoring kan daarop zeker niet een zodanige invloed hebben gehad, dat daarmee de spreiding in de tabel wordt verklaard. Noch voor de jongens, noch voor de meisjes, noch ook voor hun gecombineerde gemiddelde, wordt de toevalsverdeling benaderd.

We kunnen er daarom van uitgaan, dat de hier gegeven verschillen in de standaardverdeling essentieel zijn. De items van de verschillende schalen hebben blijkbaar een zeer uiteenlopende aantrekkingskracht. Dit zou kunnen worden veroorzaakt, doordat de individuele items een verschillend appèl uitoefenen, of omdat de mate van toewending tot de belangstellingsgebieden verschilt. Ongetwijfeld is de eerstgenoemde factor werkzaam: bij tabellering van de frequentieverdeling der itemscore blijkt duidelijk, dat de ene item gemiddeld veel vaker wordt gekozen dan de andere, ook dan b.v. een andere item uit dezelfde schaal. Maar het is bij een combinatie van 18 items in een schaal buitengewoon onwaarschijnlijk, dat deze individuele voorkeursverschillen elkaar in aanzienlijke mate zullen versterken. In hoofdzaak zullen de geconstateerde verschillen in de standaardverdeling wel teruggaan op verschillen in toewending tot de belangstellingsgebieden.

Bij de jongens blijken dan de technische handarbeid en de agrarische activiteiten sterk in trek te zijn, terwijl ook de exacte, meer theoretisch gerichte bezigheden in tel zijn. Er is daarentegen een zeer geringe toewending tot de administratieve en de sociale activiteiten. In mindere mate geldt dit ook voor de literair-geesteswetenschappelijke sector.

Bij de meisjes zijn de verschillen veel extremer dan bij de jongens, ener-

zijds door de zeer lage scores voor technische handarbeid, techniek en natuurwetenschap, administratie en handel, anderzijds door de zeer hoge scores voor ambachtelijke vormgeving en vooral sociaal werk en opvoeding. In de praktijk van de beroepskeuzevoorlichting kan men constateren, dat inderdaad een bijzonder groot aantal meisjes op de overgang van lager naar voortgezet onderwijs beroepswensen in de laatstgenoemde sector hebben. We mogen daarom wel aannemen, dat de zeer hoge scores voor dit gebied inderdaad een reële betekenis hebben. Zij zijn niet alleen een artefact van een test, die dwingt tot een keuze in een situatie, waarin een aantal der geboden mogelijkheden bij voorbaat als onaantrekkelijk kunnen worden gekenmerkt, maar de uiting van een reël aanwezige krachtige drijfveer.

Vergeleken bij de beide genoemde zeer hoge scores, vallen de ook iets aan de hoge kant liggende scores voor agrarische arbeid en literair-geesteswetenschappelijk werk nauwelijks op.

De hier geconstateerde verschillen zijn voor degene, die geregeld kinderen van de betreffende leeftijdsgroep moet interviewen, in verband met b.v. hun schoolkeuze, geen openbaring. Zij liggen geheel in de lijn der verwachtingen. Juist daarom zijn zij echter waardevol, omdat hiermee de verankering van de testgegevens in de realiteit mogelijk wordt, waardoor we ook aan de bevindingen in situaties, waar we a priori weinig of niets over weten, waarde kunnen gaan hechten.

We kunnen nu vervolgens de reacties van jongens en meisjes op de test onderling vergelijken. We hebben dit op twee wijzen gedaan.

In de eerste plaats is op de benaderde mediaan van het totale materiaal (dus jongens en meisjes tezamen) de significantie van het verschil in de sexeverdeling berekend. Onder de benaderde mediaan wordt in dit geval verstaan de scheidingslijn tussen twee scores, die zo dicht mogelijk bij de werkelijke mediaan ligt. De significantiebepaling geschiedde in eerste instantie door middel van een χ^2 -toets; aangezien deze echter zeer hoge waarden voor de χ^2 -functie en dus ook voor het significantieniveau opleverde, is de uitkomst van de berekening daarna omgezet in een *phi*-coëfficiënt (punttetrachorische correlatie), volgens een bekende formule (57, pag. 340). De coëfficiënt is zo gekozen, dat een positieve waarde wijst op een hogere mediale score bij de jongens.

In de tweede plaats zijn de verdelingen meer in detail vergeleken, door inspectie van de decielenverdeling. De uitkomsten van deze controle zullen schaalsgewijs worden gepresenteerd.

Technische handarbeid (TH). Het ligt geheel in de lijn der verwachtingen, dat de jongens een veel grotere voorkeur voor de items uit deze schaal tonen dan de meisjes. Nog geen 10% van de meisjes bereikt een score van 10 punten, een waarde, die door 80% van de jongens wordt bereikt of overtroffen. Er is daardoor een zeer hoge correlatie tussen sexe en voorkeur voor deze schaal ($\phi = +.71$). Incidenteel komen echter ook

bij meisjes wel hoge scores op deze schaal voor, de hoogste waarde door een meisje bereikt — een uitbijter — bedroeg 66.

Ambachtelijke vormgeving (AV). Hier doet zich het omgekeerde verschijnsel voor. De meisjes behalen nu duidelijk hogere scores, al is het verschil niet zo geprononceerd als in het eerste geval. Van de meisjes behaalt 80% een score, hoger dan 20 punten, terwijl slechts 30% van de jongens dit niveau bereikt ($\phi = -.48$).

Techniek en natuurwetenschap (TN). Deze schaal heeft een hoge positieve correlatie met TH, het ligt dus wel voor de hand en ook overigens in de lijn van de verwachting, dat de jongens hier meer voorkeur zullen tonen dan de meisjes. Van de jongens behaalt 80% een score van meer dan 8 punten, tegen nog geen 20% van de meisjes. De overlapping is dus iets groter dan bij de eerste schaal, hetgeen ook in de iets lagere correlatie ($\phi = +.65$) tot uiting komt.

Voedselbereiding (VB). Deze schaal bevat een aantal items, die waarschijnlijk meer in de manlijke, dan in de vrouwelijke belangstellingssfeer liggen (b.v. „een wijnhandel leiden”). Wellicht mede hierdoor is het sexeverschil minder groot dan men zou verwachten: over het grootste gedeelte van de schaal liggen de normen van de jongens twee decielen verschoven ten opzichte van die van de meisjes. Van de meisjes behaalt 60% een score van tenminste 17 punten, terwijl 60% der jongens een score van niet meer dan 17 punten heeft. De overlapping bedraagt dus 40% ($\phi = -.19$).

Agrarische arbeid (AA). Van de negen oorspronkelijke gebieden is dit de schaal met de kleinste sexeverschillen. Uit tabel 8.2 blijkt, dat er in de lage scores geen verschil is tussen jongens en meisjes, maar dat in de richting van de hoge scores het evenwicht verstoord is. Dit zou kunnen samenhangen met het feit, dat een aantal items betrekking hebben op lichamelijk vrij zware arbeid, die voor de meisjes om deze reden weinig aantrekkelijk is. In de hogere regionen bedraagt het verschil ongeveer twee decielen: 40% van de jongens behaalt een score van 27 of hoger, terwijl 20% van de meisjes 28 of meer punten behaalt ($\phi = +.13$).

Commerciële arbeid (HA). Over het gehele gebied behalen de jongens in deze sector hogere scores dan de meisjes: de schalen zijn ongeveer drie decielen ten opzichte van elkaar verschoven, de overlapping is ca. 35%. De correlatie bedraagt $\phi = +.34$.

Administratief werk (AD). Het sexeverschil heeft hier weer niet veel praktische betekenis, er is een betrekkelijk grote mate van overlapping ($\phi = +.15$). In de lagere regionen zijn de scores twee decielen ten opzichte van elkaar verschoven, bij de hogere waarden bedraagt het verschil ongeveer anderhalf deciel.

Literaire en geesteswetenschappelijke arbeid (LG). Hier bestaan vrij grote sexeverschillen in de voorkeur, waarbij ditmaal de waarden van de meisjes over de gehele lijn hoger liggen. Het verschil is het meest gepro-

nonceerd in het lage gedeelte van de frequentieverdeling: 70% van de jongens behaalt niet meer dan ten hoogste 17 punten, terwijl 60% van de meisjes 17 of meer punten behaalt. Aan de hoge zijde slinkt het verschil tot minder dan één deciel ($\phi = -.33$).

Sociaal werk en opvoeding (SO). De sexeverschillen zijn hier maximaal. Slechts 10% van de meisjes behaalt een score van minder dan 20 punten, terwijl ruim 80% van de jongens beneden deze waarde blijft ($\phi = -.73$).

Op de beide *niveauschalen* hebben de sexeverschillen zeer weinig te betekenen. Ten aanzien van de L-schaal bestaan er alleen in de extreme waarden kleine verschillen: de spreiding bij de meisjes is iets kleiner dan bij de jongens, maar het verschil is niet significant. Ten aanzien van de M-schaal treffen we in de laagste helft geen sexeverschillen aan, bij de hogere waarden krijgen de jongens een kleine voorsprong, die tot ongeveer één deciel oploopt. Dit weinig betekende verschil is interessant, omdat bij het vooronderzoek in verband met de constructie van de M-schaal wel een sexeverschil werd aangetroffen, dat toen over de gehele lijn ongeveer één kwartiel (twee-en-een-half deciel) bedroeg. Mogelijk is dit een gevolg van het feit, dat de eerste steekproef op andere wijze was samengesteld dan het ijkingsmateriaal, in ieder geval vormt deze bevinding een bevestiging van de stelling, dat kruisvalidatie noodzakelijk is.

De grote sexeverschillen op diverse schalen wijzen er op, dat het mogelijk moet zijn, met behulp van de test een *masculiniteit-feminiteits-index* te bepalen, hetzij op grond van de reacties op speciaal daartoe gekozen items, hetzij door een gewogen combinatie van schaalscores. In het kader van deze studie hebben wij de constructie en toetsing van een dergelijke schaal echter nog niet uitgevoerd.

HOOFDSTUK IX

INTERESSE EN SCHOOLKEUZE

9.1. *Inleiding*

Bij het verlaten van de lagere school moet het kind kiezen tussen een groot aantal richtingen. Deze keuze wordt ten dele op grond van de belangstelling gedaan, daarnaast spelen natuurlijk tal van andere factoren, voornamelijk van psychologische, schoolorganisatorische en maatschappelijke aard, een rol. In verband met ons onderzoek is het van belang na te gaan, in hoeverre de verschillen in keuze overeenstemmen met verschillen in reactie op de beroepeninteresseset.

Het ijkingsmateriaal biedt een goede basis, de hier eventueel bestaande relaties op te sporen. Bij het inzenden van het testmateriaal is door de onderwijzers vrijwel steeds tevens een opgave gedaan van de bestemming van hun leerlingen. Deze opgave heeft betrekking op de situatie zeer kort voor het einde van het schooljaar, in hoofdzaak zullen de gegevens dus wel overeenstemmen met de werkelijk gedane keuze. Om praktische redenen hebben wij er van afgezien, door navraag in het volgende jaar te onderzoeken, in hoeverre de opgaven nog wijzigingen behoeften. Het tijdstip van aanmelding voor de meeste schooltypen was echter ten tijde van het onderzoek reeds achter de rug, gewoonlijk was ook de uitslag van de toelatingsexamens, die de kinderen eventueel moesten afleggen reeds bekend, of althans te voorzien. In hoofdzaak mogen we dus de door ons gebruikte gegevens wel representatief achten voor de schoolkeuze.

Hoewel de concrete variatie in keuzemogelijkheden zeer groot is, kunnen we een vereenvoudiging aanbrengen, door een aantal vormen van voortgezet onderwijs tezamen te nemen. We brengen b.v. geen onderscheidingen aan binnen het v.h.m.o., we rekenen m.u.l.o. en handelsonderwijs tot één categorie en we vormen één groep lager technisch onderwijs. Al deze categorieën omvatten aanzienlijke aantallen leerlingen. Vrij grote groepen vormen ook nog degenen, die naar het nijverheidsonderwijs voor meisjes gaan, degenen die het v.g.l.o. kiezen, die rechtstreeks de praktijk in gaan en tenslotte degenen, die een brugklasse of andere overgangsinstelling gaan bezoeken, met de bedoeling, zo mogelijk het volgende jaar een toelatingsexamen voor een v.h.m.o. school af te leggen. Bij de codering van de gegevens uit het ijkingsmateriaal was ook nog de mogelijkheid geschapen, enkele andere keuzen aan te geven, maar deze bleken numeriek te weinig voor te komen, om voor bewerking in aanmerking te komen.

In het ijkingsmateriaal kwamen voor: 37 meisjes en 65 jongens met bestemming v.h.m.o., 164 meisjes en 153 jongens die naar m.u.l.o. of handelsdagonderwijs zouden gaan; 153 meisjes die naar huishoud- of industrieschool en 185 jongens, die naar een lagere technische school zouden vertrekken; 4 meisjes en 17 jongens bestemd voor verschillende andere vormen van vakonderwijs; 80 meisjes en 50 jongens met bestemming v.g.l.o., 21 meisjes en 29 jongens, die de praktijk in gingen; 20 meisjes en 25 jongens, die een brugklasse zouden gaan bezoeken; 5 meisjes en 5 jongens met onbekende bestemming en tenslotte 2 meisjes en 3 jongens, die als „diversen” zijn gerubriceerd. Onder deze laatste bevonden zich de zeer sporadische doubleergevallen.

In het volgende zullen wij voor de genoemde groepen nagaan, hoe de verdeling van de scores op de verschillende schalen is. De analyse zal voor de beide sexen gescheiden gebeuren. Aanvullend zal nog een analyse worden gerapporteerd, die betrekking heeft op een aantal zesde klassers, die een agrarisch beroep wensten te kiezen. In het ijkingsmateriaal werd de wens daartoe opvallend weinig gerapporteerd, vermoedelijk zijn de meeste aspirant-agrariërs door de onderwijzers geplaatst in de categorieën v.g.l.o. of „de praktijk in” en daardoor voor ons onherkenbaar geworden. Teneinde toch een indruk te krijgen van de karakteristiek dergenen, die de agrarische sector kiezen, zijn uit het ijkingsmateriaal de gevallen gelicht, waarin een agrarische bestemming werd aangegeven. Deze zijn gecombineerd met een aantal gevallen uit een ander onderzoek van zesde klassers, dat hier verder niet zal worden besproken. Op deze wijze werd de beschikking verkregen over 29 testformulieren van jongens, die een agrarisch beroep zouden willen kiezen. Het aantal meisjes met soortgelijke ambities, waarover gegevens beschikbaar waren, was zo gering, dat van een bewerking is afgezien.

Voor de karakteristiek van de belangstelling zal zowel in dit als in de volgende hoofdstukken gebruik worden gemaakt van een netwerk, dat ontleend is aan de in tabel 8.2 weergegeven decielenverdeling. Voor iedere onderzochte groep is per schaal een cumulatieve frequentieverdeling gemaakt, met als grenslijnen de ruwe scores, die juist beneden de opeenvolgende decielen liggen. Men kan dus uit de tabel aflezen, hoeveel kinderen een score hadden, die *lager* is dan de opeenvolgende decielwaarden. In de rij D₅ staat b.v. vermeld, hoeveel kinderen uit de onderzochte categorie een score hadden, die lager is dan de mediaan van de gehele sexegroep uit het ijkingsmateriaal. Door cursivering is de ligging van de mediaan van de onderzochte groep aangegeven. De ligging van deze mediaan ten opzichte van die van het ijkingsmateriaal geeft een eerste aanduiding van het feit, dat het gemiddelde kind uit de onderzochte categorie eventueel op een bepaalde schaal een opvallend hoge of lage score behaalt. In de laatste rij van de tabel zijn de aantallen kinderen vermeld, die een score op of boven het negende deciel behaalden. De som van de in de

beide laatste rijen van alle tabellen vermelde aantallen, is per kolom gelijk aan het totale aantal kinderen in de betreffende groep. Dit aantal (N) is steeds ook in de kop van de tabel vermeld.

De letteraanduidingen boven de kolommen zijn die van de elf belangstellingsschalen. De verklaring van de afkortingen is te vinden in hoofdstuk III, terwijl zij bovendien vermeld is op een bij het boek gevoegde bladwijzer.

9.2. Bestemming v.h.m.o.

De tabellen 9.1 en 9.2 geven de frequentieverdeling van de schaa scores van respectievelijk meisjes en jongens in vergelijking met de ijkingsgroep.

TABEL 9.1. Belangstellingskarakteristiek van meisjes, die v.h.m.o. kiezen (N = 37)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	7	—	8	1	10	7	3	5	7	1
D ₂	—	10	3	11	6	13	14	5	9	13	1
D ₃	8	12	—	18	6	22	21	8	14	17	6
D ₄	—	17	6	20	9	22	23	10	16	21	6
D ₅	16	21	9	27	12	27	24	12	18	25	9
D ₆	26	23	13	29	13	29	27	15	23	29	12
D ₇	27	27	18	30	19	32	29	16	25	32	14
D ₈	31	30	22	31	20	34	30	18	28	36	17
D ₉	35	33	29	35	27	35	33	26	34	37	22
≥ D ₉	2	4	8	2	10	2	4	11	3	0	15

Opmerking: In de kolommen TH en TN zijn bij de meisjes enkele grenslijnen niet gebruikt, omdat de decielverdeling van het ijkingsmateriaal op de betreffende plaatsen geen onderscheiding toelaat.

Het teken < (kleiner dan) is in alle tabellen uitsluitend in de eerste rij aangegeven. Het geldt echter ook voor de overige rijen, met uitzondering van de laatste.

TABEL 9.2. Belangstellingskarakteristiek van jongens, die v.h.m.o. kiezen (N = 65).

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	18	7	4	4	13	8	6	4	3	13	0
D ₂	30	13	8	9	19	14	11	4	9	29	1
D ₃	40	19	13	19	29	18	14	9	24	36	4
D ₄	45	29	16	23	34	23	17	12	29	44	7
D ₅	49	34	23	30	38	27	26	20	32	47	13
D ₆	53	43	29	32	45	34	31	24	43	54	18
D ₇	56	45	35	38	51	41	37	31	45	55	26
D ₈	59	52	43	46	56	47	41	35	52	58	32
D ₉	61	58	49	53	61	54	48	46	58	62	46
≥ D ₉	4	7	16	12	4	11	17	19	7	3	19

Een beschouwing van deze beide tabellen toont in de eerste plaats duidelijk aan, dat de kinderen, die naar het v.h.m.o. gaan, in vergelijking met de andere zesde klassers nog lang geen homogene groep vormen. Desondanks blijkt de verdeling van hun scores aanmerkelijk af te wijken van die van de gehele groep. Daarbij vertonen de meisjes nog weer een ander beeld dan de jongens.

Bij de meisjes vinden we hoge scores in de beide „theoretische” sectoren TN en LG en nog meer geprononceerd in de samengestelde niveauschaal M, die zoals bekend voornamelijk uit items uit de beide eerstgenoemde schalen bestaat. Verder ligt ook de gemiddelde score in de agrarische schaal betrekkelijk hoog. We hebben dit ook reeds opgemerkt bij de constructie van de gewogen MO-schaal, die gebaseerd is op de tests van andere kinderen, dan hier worden bestudeerd. Toen is gebleken, dat de voorkeur van deze meisjes voornamelijk uitgaat naar agrarische bezigheden, die traditioneel enigszins het stempel van „luxe” dragen.

Aan de negatieve kant vinden we vele lage scores in de commerciële en administratieve sectoren en in de voedselbereiding. De L-schaal heeft een enigszins laag gemiddelde, maar differentieert lang niet zo goed als de M-schaal. Meisjes die v.h.m.o. kiezen onderscheiden zich van het gemiddelde meisje dus door een geprononceerde intellectuele belangstelling, ook in de exacte richting, een voorkeur voor statusgebonden agrarische activiteiten en een afwending van handel, kantoorwerk en voedselbereiding. Bij de bespreking van het onderzoek van leerlingen van het v.h.m.o. zal gelegenheid bestaan, hierover nog enkele aanvullende opmerkingen te maken. Over de predictieve waarde van het gesignaleerde voorkeurspatroon is daarmee nog niets gezegd; d.w.z. we weten niet of meisjes, die het typische beeld het beste benaderen, ook meer kans van slagen hebben, dan overigens gelijkbegaafde meisjes, die daarvan afwijken. Ook dit zal eerst in het kader van het onderzoek van v.h.m.o.-leerlingen nader bestudeerd kunnen worden.

Bij de jongens treffen we aan de hoge kant eveneens de schalen TN, LG en M aan. Betrekkelijk hoog is hier voorts de score op de AD-schaal, die bij de meisjes weinig in trek was. Aan de negatieve kant vinden we een zeer lage gemiddelde score voor de technische handarbeid en ook op de L-schaal. Voorts blijkt er gemiddeld ook weinig affiniteit tot de agrarische arbeid te bestaan. In al deze gevallen zou een afkeer van lichamelijke-zware arbeid een rol kunnen spelen.

Zowel bij jongens als bij meisjes blijkt de M-schaal een goede differentiërende werking te hebben: ongeveer de helft van alle v.h.m.o.-candidaten behaalt een score, die in het totale materiaal (met inbegrip van de ruim 10% aspirant middelbare scholieren) slechts door 20% wordt bereikt. Bij jongens heeft de L-schaal een analoge differentiërende werking, terwijl dit bij meisjes in minder sterke mate het geval is. Ten aanzien van de niveauschalen geldt dus voor jongens, dat het v.h.m.o.-patroon wordt gekenmerkt

door een hoge M- en een lage L-score en voor de meisjes door een hoge M- en een relatief lage L-score.

9.3. Bestemming m.u.l.o. of handelsdagschool

Analoog aan de beide voorgaande tabellen geven tabel 9.3 en 9.4 de verdeling van de scores van meisjes respectievelijk jongens, die m.u.l.o. of aanverwante richtingen kiezen.

TABEL 9.3. Belangstellingskarakteristiek van meisjes, die m.u.l.o. kiezen (N = 164)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	9	—	16	16	11	12	10	21	15	10
D ₂	—	27	9	32	27	20	33	18	35	34	18
D ₃	38	42	—	58	48	42	46	32	52	46	32
D ₄	—	61	54	66	69	60	55	45	66	57	40
D ₅	73	73	77	90	79	80	79	57	80	67	60
D ₆	100	94	92	97	104	95	88	76	95	95	80
D ₇	123	109	102	119	118	113	104	90	112	111	88
D ₈	144	129	126	131	133	134	122	114	135	125	111
D ₉	156	142	148	147	149	146	146	141	145	145	142
≥ D ₉	8	22	16	17	15	18	18	23	19	19	22

Het beeld, dat uit deze tabel naar voren komt, is uiterst neutraal. De meisjes, die m.u.l.o. kiezen onderscheiden zich in belangstellingskarakteristiek nauwelijks van het gemiddelde. We vinden alleen een aanduiding van een iets meer dan gemiddelde belangstelling voor de literair-geesteswetenschappelijke sector en — daarmee ongetwijfeld samenhangend — een wat meer dan gemiddelde score op de M-schaal. Ter verklaring van deze neutraliteit zouden wij willen wijzen op het speciale karakter van het m.u.l.o., dat een veel gemêleerder groep leerlingen trekt, dan de meeste andere onderwijsvormen. Enerzijds treft men er vrij veel leerlingen aan, die ook een school voor v.h.m.o. hadden kunnen kiezen, anderzijds zijn er zeer veel leerlingen, die reeds spoedig blijken geven, op dit schooltype niet thuis te horen en dan naar een nijverheidsschool verhuizen of de praktijk in gaan. Niet zelden wordt het m.u.l.o. zelfs opzettelijk in plaats van v.g.l.o. als eindonderwijs gekozen, door leerlingen, die niet de bedoeling hebben een diploma te behalen. Het verloop op dit schooltype is zoals bekend nog aanzienlijk groter, dan op de overige scholen voor voortgezet onderwijs. Onder deze omstandigheden mag men nauwelijks verwachten, dat het groepsbeeld een zeer geprononceerd karakter zal hebben.

Het beeld bij de jongens verschilt in dit geval slechts zeer weinig van dat van de meisjes. De „top” in de LG-sector is verdwenen en aangezien ook in de TN-richting geen hoog gemiddelde wordt bereikt, is de waarde van de M-score niet meer, dan gemiddeld te verwachten. Als nieuwe verschijning zien we bij de jongens frequent een lage score in de sector van de

TABEL 9.4. Belangstellingskarakteristiek van jongens, die m.u.l.o. kiezen (N = 153)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	22	10	20	14	21	12	15	11	7	15	11
D ₂	43	29	29	32	36	24	29	27	29	33	29
D ₃	59	40	47	48	54	40	38	45	43	44	41
D ₄	81	64	67	64	61	60	59	57	62	53	60
D ₅	97	76	80	80	75	64	74	72	74	66	70
D ₆	112	94	96	90	85	79	82	83	95	95	91
D ₇	119	108	111	107	98	99	94	101	104	110	102
D ₈	130	120	124	118	107	110	106	112	116	130	121
D ₉	139	128	137	135	129	130	127	138	135	142	138
≥ D ₉	14	25	16	18	24	23	26	15	18	11	15

technische handarbeid, analoog dus aan de soortgelijk lage score bij de v.h.m.o.-jongensgroep. Dat dit kenmerk bij de meisjes niet opvalt, komt voort uit het feit, dat bij de meisjes in het algemeen de score op de TH-schaal bijzonder laag is. Bij een gemiddelde score van 4 punten en een mediaan van 2 blijft er weinig marge over voor variatie naar beneden. Statistisch hebben echter ook de Mulo-meisjes een significant lage TH-score (vgl. bijlage I). Op de L-schaal treffen we bij de m.u.l.o.-jongens geen laag gemiddelde aan, wel valt op dat het aantal zeer hoge scores aan de lage kant is: 23 jongens hebben een score op het niveau D₈ of hoger, bij een verwacht aantal van ruim 30.

De „gemiddelde m.u.l.o.-leerling” zoals hij uit de statistische belangstellingskarakteristiek naar voren komt, is dus een kleurloze figuur. Zoals gezegd kan dit voortkomen uit het feit, dat deze opleiding een zeer heterogene groep leerlingen aantrekt. Een consequentie is echter wel, dat we op grond van synchrone gegevens geen groep van naar belangstelling specifieke m.u.l.o.-leerlingen kunnen vormen. In principe is desondanks de mogelijkheid niet uitgesloten, dat de belangstellingskarakteristiek voor deze onderwijsvorm predictieve waarde heeft, namelijk als zou blijken, dat kinderen met een bepaald belangstellingstype een significant betere of slechtere kans van slagen zouden hebben dan andere leerlingen. Dit vraagt echter een nader onderzoek, dat in het kader van deze studie niet is verricht.

9.4. Nijverheidsonderwijs, v.g.l.o. en praktijk bij meisjes

Bij de thans te bespreken richtingen is er dikwijls groot verschil tussen de situatie waarin de jongens respectievelijk de meisjes worden geplaatst. Daarom zullen nu de beide sexen niet in dezelfde paragraaf worden behandeld. Daarentegen is er wel aanleiding een aantal richtingen per sexe met elkaar te vergelijken. In de wereld van de beroepskeuzevoorlichting bestaat namelijk dikwijls de neiging, te veronderstellen, dat er naar type betrekkelijk weinig verschil bestaat tussen de meisjes die na de zesde klas

naar een industrie- of huishoudschool gaan en degenen die v.g.l.o. gaan bezoeken of direct de praktijk in gaan. Zo er al verschillen bestaan veronderstelt men, dat deze vooral in de sector van de intellectuele begaafdheid en eventueel van het maatschappelijk milieu zijn gelegen. Het is daarom interessant na te gaan of er naar belangstellingsrichting inderdaad betrekkelijk weinig verschil tussen deze categorieën meisjes bestaat. De tabellen 9.5, 9.6 en 9.7, die betrekking hebben op meisjes, die respectievelijk naar het n.o. het v.g.l.o. of direct naar de praktijk gaan, vormen de basis voor de vergelijking.

TABEL 9.5. Belangstellingskarakteristiek van meisjes, die n.o. kiezen (N = 153)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	16	—	9	14	6	5	11	7	5	9
D ₂	—	33	10	14	30	14	15	33	15	13	24
D ₃	27	44	—	28	51	33	34	45	23	21	43
D ₄	—	55	40	35	59	42	55	76	34	35	74
D ₅	57	69	72	61	78	72	69	89	51	54	93
D ₆	83	90	92	70	83	85	95	97	71	72	106
D ₇	99	111	125	87	111	105	118	113	90	84	120
D ₈	118	120	138	110	125	121	130	134	107	95	132
D ₉	137	140	144	125	142	135	143	142	131	123	144
≥ D ₉	16	13	9	28	11	18	10	11	22	30	9

De huishoudschoolgroep vertoont dus een duidelijker karakteristiek dan de groep van de m.u.l.o.-meisjes. Er zijn geen opvallend lage gemiddelde scores, maar wel zijn er enkele sectoren, waar de gemiddelde score vrij hoog ligt: voedselbereiding, sociaal werk en opvoeding. Ook de L-score komt boven het gemiddelde uit. De pluswaarden liggen in de sectoren, waar men deze ook op grond van de ervaring zou verwachten.

TABEL 9.6. Belangstellingskarakteristiek van meisjes, die v.g.l.o. kiezen (N = 80)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	3	—	2	4	2	0	7	8	1	9
D ₂	—	8	8	5	8	4	0	17	22	10	13
D ₃	3	16	—	18	20	9	0	21	40	17	20
D ₄	—	22	16	23	28	12	13	29	50	24	27
D ₅	8	31	27	36	41	19	19	37	55	30	33
D ₆	17	42	31	44	47	27	29	51	62	47	46
D ₇	27	50	40	52	62	37	37	61	64	52	52
D ₈	32	55	50	63	67	51	51	71	70	60	70
D ₉	50	73	63	71	75	71	67	78	74	69	79
≥ D ₉	30	7	17	9	5	9	13	2	6	11	1

We zien uit tabel 9.6, dat het beeld, dat de v.g.l.o.-meisjes vertonen, zeer sterk afwijkt van dat van de huishoudschoolgroep. Natuurlijk zullen er onder degenen, die v.g.l.o. kiezen wel een aantal zijn, die ook op de huishoudschool hadden gepast, maar er is blijkbaar een zeer grote categorie v.g.l.o.-leerlingen, die een totaal anders gerichte belangstelling heeft. Dit blijkt wel uit de opvallend hoge score in de sector van de *technische handarbeid*. We stuiten hier op een blijkbaar toch niet zeer kleine categorie vrouwen, die affiniteit hebben tot de praktische uitoefening van een technisch beroep, maar niet de mogelijkheid in de richting van hun interesses een opleiding te volgen.

Mede gezien buitenlandse ervaringen wijst dit op een leemte in onze onderwijsvoorzieningen. Het zou gewenst zijn als, voorlopig bij wijze van proef, b.v. in enkele grote steden de mogelijkheid voor meisjes werd geopend, een praktisch-technisch geschoold beroep te leren.

Onder de gegeven omstandigheden voelen deze meisjes, die vaak weinig affiniteit tot huishoudelijk werk bezitten, zich niet aangetrokken tot de keuze van een huishoudschool, zodat het v.g.l.o. als enig alternatief overblijft.

Een andere categorie, die wellicht ook niet al te veel affiniteit tot huishoudelijk werk heeft, wordt gevormd door degenen, die een voorkeur voor administratief of commercieel werk hebben; op de betreffende schalen wordt eveneens een hoge gemiddelde score bereikt. Daarentegen is het gemiddelde op de SO-schaal uitgesproken laag, ook op dit punt gedraagt de v.g.l.o.-groep zich geheel anders dan de huishoudschoolgroep. Vermelding verdient tenslotte nog de TN-score, die iets boven het gemiddelde ligt. Het kan zijn, dat dit een correlaat is van de belangstelling in de TH-sector; maar denkbaar is ook, dat zich in de groep enkele meisjes bevinden, die in feite een brugklasse voor het v.h.m.o. volgen en die met hun relatief hoge scores in de sector van techniek en natuurwetenschap, deze kleine top hebben veroorzaakt.

TABEL 9.7. Karakteristiek van meisjes, die geen voortgezet onderwijs volgen (N=21)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	1	—	1	1	0	0	6	1	1	5
D ₂	—	4	5	1	4	0	1	12	4	1	11
D ₃	0	8	—	1	4	1	2	13	5	1	13
D ₄	—	11	10	3	6	2	3	15	11	3	13
D ₅	3	13	11	6	10	4	6	16	13	3	18
D ₆	5	13	12	8	12	6	10	16	14	4	18
D ₇	7	13	15	13	12	8	12	18	16	5	18
D ₈	10	16	18	14	14	12	16	20	18	10	19
D ₉	15	18	18	16	18	15	17	21	19	13	20
≥ D ₉	6	3	3	5	3	6	4	0	2	8	1

Het aantal meisjes, dat geen verder dagonderwijs volgt is betrekkelijk klein, tabel 9.7 biedt daardoor niet veel mogelijkheden tot interpretatie. In deze groep zullen zich vele betrekkelijk zwakbegaafden bevinden, die de capaciteiten tot het leren van een beroep op geschoold niveau missen. Vandaar waarschijnlijk de lage score op de „intellectuele” schalen, in het bijzonder LG en M. Het karakteristieke patroon is overigens verwant aan dat van de v.g.l.o.-candidaten, vooral in de relatief hoge TH en HA-scores. Maar de karakteristiek is ditmaal minder geprononceerd, als regel staat deze groep tussen de v.g.l.o.-categorie en degenen, die huishoudschool kiezen.

Een eigen kenmerk is de als regel hoge L-score. Veronderstellenderwijs kan de mogelijkheid worden geopperd, dat meisjes op eenvoudig begaafdheidsniveau gemiddeld een masculiener type vertonen, dan de beter begaafden.

9.5. Nijverheidsonderwijs, v.g.l.o. en praktijk bij jongens

Analoog aan de drie tabellen uit de voorgaande paragraaf geven de tabellen 9.8, 9.9 en 9.10 een overzicht van de belangstellingskarakteristiek van de drie corresponderende jongensgroepen.

TABEL 9.8. Belangstellingskarakteristiek van jongens, die l.t.o. kiezen (N = 185)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	2	11	7	14	10	10	12	8	12	3	11
D ₂	12	26	24	32	24	21	20	35	28	13	32
D ₃	20	42	39	47	40	35	50	51	45	27	50
D ₄	34	61	63	58	68	58	73	68	66	34	75
D ₅	55	87	79	92	90	66	85	92	87	52	95
D ₆	70	95	99	104	115	109	115	114	99	77	117
D ₇	98	124	123	116	137	135	131	143	125	91	143
D ₈	129	141	149	147	159	155	154	163	149	120	160
D ₉	146	168	173	167	174	170	181	175	167	151	176
≥ D ₉	39	17	12	18	11	15	4	10	18	34	9

De uitkomsten van de analyse liggen wel in de lijn van de verwachtingen: een hoge gemiddelde L-score, een betrekkelijk hoge TH-score en een tekort aan hoge waarden in de AD-richting en op de M-schaal. Veel meer differentiatie is hier niet te verwachten, vooral omdat de lagere technische school verscheidene richtingen omvat. De karakteristiek van deze richtingen loopt in sommige gevallen sterk uiteen, waardoor een statistische nivellering optreedt. Bij het ijkingsmateriaal was het niet mogelijk, over de details van de keuze van het lager technisch onderwijs gegevens te krijgen, enerzijds omdat vele onderwijzers van de lagere school over deze vorm van voortgezet onderwijs geen voldoende heldere voorstellingen hebben, anderzijds omdat tegenwoordig het lager technisch onderwijs aan vele

scholen begint met een voorbereidend jaar en pas in het tweede jaar een eerste differentiatie in vakrichtingen volgt. We zullen echter in een afzonderlijk hoofdstuk nog nader op de verschillen in belangstellingskarakteristiek binnen de lagere technisch school ingaan.

TABEL 9.9. Karakteristiek van jongens, die v.g.l.o. kiezen (N = 50)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	0	3	5	0	4	4	5	4	1	6	6
D ₂	2	6	11	3	6	9	7	9	6	9	12
D ₃	10	8	18	5	8	16	9	14	9	13	14
D ₄	15	11	21	13	17	22	19	22	17	18	18
D ₅	25	17	22	13	20	25	22	26	18	22	23
D ₆	31	22	26	16	28	32	27	29	24	28	27
D ₇	41	29	34	26	30	38	36	33	31	34	33
D ₈	44	37	40	37	38	42	44	41	39	39	41
D ₉	48	43	45	43	46	48	49	44	46	43	44
≥ D ₉	2	7	5	7	4	2	1	6	4	7	6

Het is niet uitgesloten, dat zich ook hier in de v.g.l.o.-groep enkele kinderen bevinden, die zich voorbereiden op het toelatingsexamen van een v.h.m.o.-school. In dat geval zou van hen enige invloed op de hoogte van de TN-, LG- en M-scores kunnen zijn uitgegaan, waardoor deze niet beneden het gemiddelde zijn gekomen.

Evenals bij de meisjes blijkt ook bij de jongens de v.g.l.o.-groep een aanmerkelijk andere karakteristiek te hebben, dan de categorie, die een lagere technische school gaat bezoeken. Er zijn geen opvallend lage gemiddelde scores. Maar in drie gevallen, ambachtelijke vormgeving, voedselbereiding en sociaal werk en opvoeding, schijnt het gemiddelde betrekkelijk hoog te liggen. Deze drie schalen zijn in het algemeen bij meisjes meer in trek dan bij jongens. Men zou dit kunnen zien als een aanwijzing, dat het v.g.l.o. een meer dan gemiddelde aantrekkingskracht uitoefent op jongens van doorsnee-begaafdheid, met een wat feminiene inslag. Deze zullen zich in de meeste richtingen van de lagere technische school niet bijzonder op hun plaats voelen. Statistisch blijkt echter alleen het verschil op de AV-schaal significant. De hier geuite veronderstelling behoeft dus nader bewijs.

Gezien het kleine aantal personen laat tabel 9.10 geen vérstrekkende interpretaties toe. De karakteristiek blijkt echter in hoofdzaak verwant aan die van de v.g.l.o.-jongens, met dien verstande, dat op de beide „intellectuele” schalen TN en M de gemiddelde score wat lager is. In dit opzicht is er verwantschap met de meisjes, die geen voortgezet onderwijs volgen. Dit geldt ook wat betreft de gemiddeld hoge L-score.

Er ligt dus in het materiaal een vingerwijzing, dat de psychologische geslachtsverschillen, zoals deze in de test tot uiting komen, correleren

TABEL 9.10. Karakteristiek van jongens, die geen voortgezet onderwijs volgen
(N = 29)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	2	2	5	3	2	1	1	0	0	0	6
D ₂	5	5	10	4	3	2	2	2	1	5	10
D ₃	6	6	12	5	5	6	4	7	3	6	11
D ₄	9	7	19	7	9	10	7	11	5	7	14
D ₅	13	8	22	10	10	12	10	19	6	10	21
D ₆	17	11	26	12	14	16	15	20	9	12	23
D ₇	21	14	27	16	18	17	24	24	14	14	24
D ₈	23	18	27	22	22	24	29	26	19	15	28
D ₉	27	25	28	26	26	25	29	28	25	22	29
≥ D ₉	2	4	1	3	3	4	0	1	4	7	0

met het begaafdheidsniveau. De differentiatie blijkt het geringst te zijn bij de beide groepen, waarvan men met goede reden mag aannemen, dat het gemiddelde verstandelijke begaafdheidsniveau het laagste is, in de eerste plaats bij degenen, die v.g.l.o. gaan bezoeken en daarna bij degenen, die na de zesde klas geen voortgezet onderwijs meer volgen.

9.6. Beroepskeuze in de agrarische sector

Tabel 9.11 geeft de frequentieverdeling van de scores der 29 jongens, die een speciaal samengestelde steekproef uit degenen die een agrarisch beroep kiezen vormen.

TABEL 9.11. Belangstellingskarakteristiek van jongens, die de agrarische richting kiezen (N = 29)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	1	2	0	2	0	3	1	1	3	1	1
D ₂	4	4	6	5	1	6	2	4	5	4	8
D ₃	9	5	12	5	1	9	9	7	7	5	11
D ₄	12	6	17	10	3	11	13	10	8	6	14
D ₅	17	9	21	13	4	16	15	17	10	10	17
D ₆	22	12	26	14	6	18	18	21	15	16	22
D ₇	23	19	27	17	8	21	22	24	18	19	26
D ₈	26	21	28	22	9	28	28	27	20	24	28
D ₉	28	25	28	25	19	28	28	28	24	27	28
≥ D ₉	1	4	1	4	10	1	1	1	5	2	1

De als regel zeer hoge score op de AA-schaal wijst er in de eerste plaats op, dat de belangstellingskarakteristiek hier een aanzienlijke mate van synchrone validiteit bezit, dit temeer, omdat we een dergelijk kenmerk bij de andere jongensgroepen niet hebben aangetroffen. Bovendien is het

feit, dat juist de AA-score als regel hoog is, een bijdrage tot de begripsvaliditeit.

Naast het primaire kenmerk van de hoge AA-score, kunnen we enige verwantschap in het voorkeurspatroon met dat van de in de vorige paragraaf besproken v.g.l.o.- en praktijk-groep constateren. Met deze groepen zal de thans aan de orde zijnde categorie naar begaafdheidsniveau ook wel ongeveer overeenkomen.

9.7. *Leerlingen in een overgangsjaar*

In het materiaal bevonden zich 45 kinderen (20 meisjes, 25 jongens), van wie werd vermeld, dat zij bestemd waren t.z.t. toelatingsexamen voor een middelbare school te doen, nadat zij eerst een overgangsjaar (brugklasse, speciale v.g.l.o.-school of -klasse, speciale zesde klasse) hadden gehad. Uit de praktijk is bekend, dat er zeer verschillende redenen voor het volgen van een dergelijk jaar kunnen bestaan. In enkele plaatsen is het bezoek van een brugklasse of dergelijke regel, behalve voor zeer begaafde leerlingen, omdat de zesde klassen niet voorbereiden op het toelatingsexamen. Elders pleegt men alleen kinderen, die men nog niet rijp acht voor het middelbaar onderwijs, bij wie men twijfelt aan de geschiktheid, of die om externe redenen, b.v. ziekte, een achterstand in kennis hebben gekregen, naar een dergelijke klasse te zenden. Een gedeelte van deze kinderen komt in werkelijkheid niet op een middelbare school terecht, maar kiest na het overgangsjaar m.u.l.o. of nijverheidsonderwijs. Daarom is het gewenst, deze groep afzonderlijk van de v.h.m.o.-groep te bestuderen. De tabellen 9.12 en 9.13 geven de verdeling voor meisjes, respectievelijk jongens.

TABEL 9.12. Belangstellingskarakteristiek van meisjes, die een overgangsjaar volgen (N = 20)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	3	—	3	0	2	3	0	0	3	0
D ₂	—	4	0	3	2	7	5	3	2	3	5
D ₃	2	7	—	6	2	10	6	4	2	3	6
D ₄	—	8	3	7	2	12	11	8	7	3	8
D ₅	7	8	4	10	3	15	12	13	11	8	11
D ₆	10	11	7	12	4	18	15	16	13	14	12
D ₇	15	12	12	13	8	18	16	18	14	15	15
D ₈	16	13	15	17	8	20	18	18	16	15	17
D ₉	18	16	19	20	11	20	18	19	18	17	18
≥ D ₉	2	4	1	0	9	0	2	1	2	3	2

Gezien het kleine aantal meisjes laat de tabel niet veel gelegenheid tot het trekken van conclusies. Opvallend is echter de zeer hoge gemiddelde score op de AA-schaal, waarvan reeds eerder is opgemerkt, dat we ver-

moedelijk met een correlaat van het sociaal-economisch milieu te maken hebben. Dit suggereert, dat we bij deze groep meisjes te maken hebben met een categorie, waarin, wellicht nog meer dan in de v.h.m.o.-groep een selectie naar sociaal-economisch niveau heeft plaatsgevonden. Deze conclusie heeft voldoende waarschijnlijkheid. Voorts blijkt de HA-score aan de lage kant te zijn, maar overigens vertoont de groep in veel opzichten overeenkomst met de categorie van de m.u.l.o.-meisjes, d.w.z. een neutraal beeld, zonder specifieke voorkeursrichtingen.

TABEL 9.13. Belangstellingskarakteristiek van jongens, die een overgangsjaar volgen (N = 25)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	1	7	0	0	2	2	3	0	6	4	1
D ₂	2	12	1	3	10	6	4	3	7	10	4
D ₃	5	15	3	3	11	6	7	7	12	11	4
D ₄	9	17	4	8	14	9	8	11	15	12	4
D ₅	13	19	5	13	14	11	10	12	18	13	6
D ₆	15	19	9	17	14	12	11	12	18	16	7
D ₇	18	20	11	18	17	16	14	15	20	17	8
D ₈	20	21	12	19	19	19	19	18	22	20	13
D ₉	24	24	14	21	19	21	22	23	24	22	17
≥ D ₉	1	1	11	4	6	4	3	2	1	3	8

Ook voor de jongens geldt, dat het kleine aantal proefpersonen het trekken van verstrekkende conclusies verhindert. Opvallend hoog zijn echter de scores op de TN- en de M-schaal, terwijl de AV-score naar beneden een significante afwijking vertoont. De beide eerstgenoemde karakteristieken passen bij de kenmerken van de v.h.m.o.-jongens, bij wie echter geen lage AV-score werd geconstateerd. Wanneer we van dit laatste verschil afzien, kunnen we zeggen, dat uit de analyse de suggestie volgt, dat er geen opvallend verschil in belangstellingskarakteristiek tussen de jongens, die rechtstreeks naar het v.h.m.o. gaan en degenen, die eerst een overgangsjaar volgen, bestaat.

9.8. Samenvatting

Het voorgaande onderzoek heeft aangetoond, dat er parallel aan de differentiatie na het voortgezet onderwijs, een differentiatie in de belangstellingskarakteristiek bestaat. Ten dele liggen de gevonden verschillen in de lijn van de verwachting, voor een ander deel geven zij ook nieuwe inzichten. Mede gezien het belang voor het praktisch gebruik van de test, schijnt het daarom wel nuttig, de conclusies van het voorgaande te resumeren.

Vooropgesteld mag worden, dat in de voorgaande analyses geen poging is gedaan, alle statistisch significante verschillen op te sporen. Bij de grote

omvang van het vergelijkingsmateriaal en de eveneens vrij grote omvang van een aantal der besproken groepen, worden reeds betrekkelijk kleine verschillen — in sommige gevallen reeds verschillen van één à twee punten — statistisch significant. Voor de practijk hebben deze echter geen waarde, indien men een differentiatie van individuen beoogt. Daarom zijn verschillen van één deciel in de ligging van de mediaan van de onderzochte groep ten opzichte van de ligging van de mediaan van de vergelijkingsgroep steeds buiten beschouwing gelaten. Het vermelden van significantiewaarden heeft dan echter ook weinig zin, terwijl anderzijds de correlaties, gezien de grote spreiding binnen alle groepen toch nog zeer klein zijn. Alleen bij de groepen van kleine omvang is aan het significantieprobleem iets meer expliciet aandacht geschonken. Hier kan ook een verschil, dat b.v. juist de 5%-drempel van betrouwbaarheid bereikt, van betekenis zijn. Een overzicht van alle significante verschillen wordt in bijlage 1 gegeven.

De lage correlatiewaarden, waarvan boven sprake was, wijzen er reeds op, dat een belangstellingstest zeker niet het eerstaangewezen medium is, om een differentiatie ten behoeve van het voortgezet onderwijs door te voeren. Ongetwijfeld is b.v. een intelligentietest, in dit opzicht, als psychologisch hulpmiddel veel effectiever. Maar het feit, dat ook de belangstellingstest vermag te differentiëren en dat de correlatie tussen de schaalwaarden en het intelligentieniveau over het algemeen zeer laag is, wijst er op, dat ter aanvulling een interestest bruikbaar is. De overweging, dat de interessen bij de leeftijdsgroep in kwestie over het algemeen nog weinig stabiel zijn, vormt hiertegen geen contra-indicatie, de resultaten van de analyse wijzen immers op een — zij het bescheiden — mate van synchrone validiteit.

Wat nu de concrete uitkomsten ten aanzien van de meisjes betreft, noemen we als eerste opvallende resultaat het neutrale beeld van de m.u.l.o.-groep, dat vooral treft, indien we daartegenover de zeer geprononceerde karakteristiek van de v.h.m.o.-groep plaatsen. Ten aanzien van de m.u.l.o.-candidaten heeft de belangstellingstest practisch geen synchroon selecterende werking. Hoewel dit schooltype voortgezet algemeen vormend onderwijs geeft, heeft een groot deel der leerlingen niet meer dan een doorsnee intellectuele belangstelling.

Opvallend en tevens onverwacht is ook het grote verschil tussen degenen, die naar een huishoudschool zullen gaan enerzijds en degenen, die naar v.g.l.o. gaan of het onderwijs verlaten anderzijds. Bij eerstgenoemden komt het vrouwelijk-zorgende element in de belangstellingskarakteristiek vrij duidelijk naar voren. We zouden hierbij aansluitend de veronderstelling willen uitspreken, dat een test, waarin ook voor de zorg voor de kleding en voor de huishouding een schaal zou zijn opgenomen, ten aanzien van deze categorie leerlingen waarschijnlijk de mogelijkheid tot selectie zou vergroten. Bij de overige meisjes hebben we een verschuiving van het beeld in de richting van een voorkeur voor de „manlijke” interes-

sengebieden opgemerkt. Dit kwam het sterkste tot uiting bij degenen, die naar het v.g.l.o. gaan, zodat we concluderen, dat hier de meisjes hun toevlucht zoeken, die zich op een huishoudschool niet thuis voelen en die anderzijds niet voor voortgezet algemeen vormende scholing in aanmerking komen.

Er schijnen bij het v.g.l.o. twee tendenties in dezelfde richting te werken. Enerzijds menen we bij de meisjes van eenvoudig sociaal-economisch niveau en meestal matige intellectuele begaafdheid in het algemeen te kunnen opmerken, dat het specifiek vrouwelijk-zorgende element niet zo sterk op de voorgrond treedt. Anderzijds is er ongetwijfeld, los van het voorgaande, een categorie meisjes, die meer masculien-gerichte interesses heeft. Voor zover deze categorie deel uitmaakt van de meerderheid, die geen voortgezet algemeen vormend onderwijs na de lagere school gaat volgen, tendeert zij tot concentratie binnen het v.g.l.o. Speciaal voor deze groep hebben wij een hiaat in onze onderwijsvoorzieningen gesignaleerd.

Bij de jongens vinden we, ondanks alle verschillen in het concrete interessespatroon, in hoofdzaak een soortgelijk onderscheid tussen de groepen met verschillende bestemming. Ook hier missen de aanstaande m.u.l.o.-leerlingen een geprononceerde belangstellingskarakteristiek. Zij hebben alleen met de aanstaande v.h.m.o.-leerlingen gemeenschappelijk een vrij duidelijke afwending van de manuele, in de eerste plaats metaal-technische arbeid. Bij vergelijking van de jongens, die naar het lager technisch onderwijs gaan, met degenen die v.g.l.o. kiezen of de praktijk in gaan, vinden we het pendant van de situatie bij de meisjes. Ook hier zou de sexe-differentiatie in de betreffende subcultuur minder — of eventueel anders! — geprononceerd kunnen zijn dan in de bevolking als totaliteit.

Opnieuw is het verschil met het nijverheidsonderwijs bij de v.g.l.o.-groep opvallend. Blijkbaar komt op deze school ook een groep jongens, die op geschoold niveau een beroep met een wat meer feminiene inslag prefereert. In dit geval zijn er wel mogelijkheden om een dergelijk beroep (kleermaker, bontwerker, kapper, banketbakker, etaleur, bepaalde varianten van winkelverkoper, e.d.) in de praktijk te leren, maar er is slechts zelden een schoolopleiding voor beschikbaar.

Tenslotte heeft een analyse van de reacties van een groep agrariërs, althans ten aanzien van de AA-schaal, ook een belangrijke bijdrage tot de vaststelling van de begripsvaliditeit geleverd.

HOOFDSTUK X

BELANGSTELLING EN RICHTINGKEUZE IN HET NIJVERHEIDS- ONDERWIJS VOOR JONGENS

10.1. *Probleemstelling*

In het vorige hoofdstuk hebben we geconstateerd, dat degenen, die het lager technisch onderwijs gaan bezoeken, zich in verschillende opzichten naar belangstellingskarakteristiek onderscheiden van de overige zesde klassers. Het was echter in het kader van het daar besproken onderzoek niet mogelijk, een differentiatie te maken naar de diverse vakrichtingen binnen dat onderwijs. Toch kan wel bij voorbaat worden vermoed, dat er ook tussen de leerlingen van deze richtingen verschillen in de reacties op de B I T zullen bestaan. Weliswaar moeten we er op verdacht zijn, dat de richtingkeuze op de lagere technische school lang niet altijd wordt gemotiveerd door de belangstelling. De keuze moet worden gedaan op een ogenblik, waarop zeer vele jongens nog slechts een flauwe notie hebben over de inhoud van het onderwijsprogramma van de verschillende vakrichtingen en van de aard van de beroepswerkzaamheden, die daarop als regel aansluiten. Zeer herhaaldelijk zal daardoor de keuze vrijwel blindelings geschieden, in navolging van b.v. de keuze van een familielid of vriend, op grond van een aanwijzing van ouders of onderwijzers, of als modeverschijnsel. Bovendien hebben vele lagere technische scholen in sommige vakrichtingen te weinig plaatsen beschikbaar om alle aanbod op te nemen. Het is geen uitzondering, wanneer dan aan degenen, die niet aanstonds geplaatst kunnen worden, de gelegenheid wordt geboden, een opleiding te volgen in een minder in trek zijnde richting. Lang niet altijd heeft dus de richtingkeuze het karakter van een beslissing, die in overeenstemming met de geaardheid van de persoon is genomen. Indien er desondanks voorkeursverschillen optreden, zijn deze voor de praktijk van de differentiatie van bijzondere betekenis.

Aangezien het moeilijk is, op de lagere school, zelfs omstreeks het einde van de zesde klas, vast te stellen welke richting een jongen op de lagere technische school zal gaan kiezen, zijn wij bij het thans te beschrijven onderzoek uitgegaan van jongens, die reeds een keuze hadden gedaan, dus leerlingen van het lager technisch onderwijs. Dit heeft onvermijdelijk als gevolg, dat de groep gemiddeld wat ouder is, dan de leerlingen die in de zesde klasse voor de beslissing staan. Bovendien hebben zij een zekere mate van vormende invloed van het voortgezet onderwijs ondergaan. Deze

bezwaren moeten worden geaccepteerd, wij hebben alleen getracht ze tot een minimum terug te brengen, door leerlingen te onderzoeken, die zeer recent tot de l.t.s. waren toegelaten. De combinatie van eisen, dat de richtingkeuze reeds moet zijn geschied en dat de toelating eerst recent heeft plaatsgevonden, bracht mee, dat het onderzoek moest plaatsvinden in de situatie van een l.t.s. zonder voorbereidend jaar. De toelating tot een dergelijke school kan evenwel pas geschieden, nadat de leerling de leeftijd van 12 jaar en 8 maanden heeft bereikt. Onze proefpersonen zijn daardoor vrijwel allen tenminste 13 jaar. Vele leerlingen der lagere school moeten echter één of meermalen doubleren, zodat zij aan het einde van de zesde klas aan de genoemde leeftijdsvoorwaarde kunnen voldoen. De meeste leerlingen die in het onderzoek waren betrokken, hadden de l.t.s. dan ook rechtstreeks uit de zesde klas bereikt.

Door de medewerking van een lagere technische school met een groot aantal vakrichtingen was het mogelijk, ook een aantal opleidingen in het onderzoek te betrekken, die in ons lager technisch onderwijs slechts betrekkelijk zeldzaam vertegenwoordigd zijn: loodgieters, schilders, typografen, banketbakkers. Deze richtingen hebben in het algemeen slechts weinig leerlingen. Teneinde toch een materiaal van voldoende omvang te verzamelen, is in deze gevallen zo nodig van de hiervoor gestelde beperkende voorwaarden afgeweken en zijn ook een aantal leerlingen uit het tweede leerjaar onderzocht. In totaal zijn in het onderzoek 510 jongens betrokken:

87 bakkers	59 timmerlieden en meubelmakers
103 bankwerkers	38 schilders
75 electriciens	21 straatmakers
51 instrumentmakers	44 typografen (zetters en drukkers)
32 loodgieters	

Het zou voor de hand liggen, aan de in ons onderzoek te betrekken proefpersonen nog een extra eis te stellen, namelijk dat zij een positieve attitude ten opzichte van de door hen gekozen richting bezitten. Het blijkt echter zeer moeilijk, bij de betreffende leeftijdsgroep betrouwbare uitspraken over hun waardering van de gekozen richting te krijgen. Te dikwijls hebben momentane invloeden en extreme overwegingen invloed op de mededelingen, die de jongens hieromtrent doen. Daarom hebben wij er van afgezien, deze eis te stellen.

Van ieder der genoemde groepen zal nu de frequentieverdeling der test-scores op dezelfde wijze worden geanalyseerd, als in het voorgaande hoofdstuk met de bestemmingsgroepen is geschied.

10.2. *Bankwerkers*

Tabel 10.1 geeft de verdeling van de testcores van 103 leerlingen van het eerste leerjaar uit de richting bankwerken.

TABEL 10.1. Belangstellingskarakteristiek van leerling-bankwerkers (N = 103)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	3	15	1	13	8	9	16	2	7	2	1
D ₂	6	30	6	24	23	14	20	22	21	5	10
D ₃	15	35	15	28	37	23	42	32	31	11	22
D ₄	23	46	25	39	55	31	55	40	42	17	37
D ₅	26	60	34	48	66	39	62	50	56	25	47
D ₆	33	71	40	56	76	47	71	61	66	43	61
D ₇	42	80	59	68	85	61	76	75	73	53	76
D ₈	54	87	73	80	89	83	90	83	80	69	89
D ₉	64	96	89	96	98	90	98	89	92	90	97
≥ D ₉	39	7	14	7	5	13	5	14	11	13	6

Het beeld beantwoordt aan de verwachtingen, die men op grond van observaties over leerling-bankwerkers kan opstellen. De mediale score in de sector van de technische handarbeid ligt zeer hoog, terwijl ook de daarmee positief gecorreleerde TN-score een betrekkelijk hoog gemiddelde bereikt. De mediale waarde op de HA-schaal ligt eveneens hoog, het gemiddelde blijkt echter niet significant van de verwachte waarde af te wijken. In negatieve richting valt in de eerste plaats de afwenning van de agrarische arbeid en het administratieve werk op. Verder valt op, dat in de AV-schaal een groot aantal zeer lage scores voorkomt, hoewel overigens de verdeling niet opvallend van de standaard afwijkt. De L-score, die zowel op items uit de TH en HA-schalen als uit de VB en AV-schalen is gebaseerd, is door deze tegenstrijdige tendenties niet zo hoog, als wellicht zou worden verwacht. Desondanks ligt de score toch nog gemiddeld duidelijk hoger dan bij de vergelijkingsgroep.

10.3. Instrumentmakers

In tabel 10.2 wordt een overzicht gegeven van de scores van 51 leerling-instrumentmakers, allen uit het eerste jaar van de opleiding afkomstig.

TABEL 10.2. Belangstellingskarakteristiek van leerling-instrumentmakers (N = 51)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	0	13	0	7	10	5	6	0	2	0	0
D ₂	2	24	4	11	20	8	9	4	13	4	3
D ₃	3	28	4	18	26	18	14	9	23	7	5
D ₄	4	32	5	24	36	21	19	16	29	8	6
D ₅	4	37	6	31	40	25	25	20	31	16	9
D ₆	6	40	10	35	45	30	32	23	37	26	18
D ₇	13	45	12	41	46	40	35	28	42	32	22
D ₈	21	47	16	49	49	45	45	36	44	36	33
D ₉	32	48	29	51	51	50	49	49	47	44	41
≥ D ₉	19	3	22	0	0	1	2	2	4	7	10

Het ligt voor de hand, dat deze groep in veel opzichten verwante trekken met de bankwerkers zal vertonen. Dit blijkt vooral uit de zeer hoge scores op de TH- en de TN-schalen. Daardoor is ook de score op de M-schaal gemiddeld hoog. Anderzijds zijn de scores op de AV- en AA-schalen gemiddeld laag, terwijl op de VB-schaal het aantal zeer hoge scores aan de lage kant is. De HA- en AD-schalen vertonen ditmaal geen bijzondere kenmerken, terwijl het bijzondere van de L-score vooral ligt in het feit, dat deze een niet meer dan gemiddelde hoogte bereikt. De totaal-karakteristiek is die van een cerebraal ingestelde, op technische bewerkingen gerichte mens, in artistiek opzicht weinig creatief en misschien juist daardoor in staat tot het minutieus vervaardigen van instrumenten en gereedschappen.

10.4. *Electromonteurs*

Van de 75 electromonteurs in ons materiaal zijn er 41 leerling van het eerste en 34 van het tweede leerjaar. Tabel 10.3 geeft een overzicht van de frequentieverdeling.

TABEL 10.3. Belangstellingskarakteristiek van leerling-electromonteurs (N = 75)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	1	12	1	6	8	5	9	4	6	2	1
D ₂	1	28	2	10	24	16	15	10	14	8	7
D ₃	4	40	2	22	34	25	22	17	21	15	8
D ₄	10	48	9	35	39	29	31	25	35	19	13
D ₅	16	53	12	42	46	33	36	38	47	24	15
D ₆	23	56	15	47	53	41	53	45	50	44	21
D ₇	28	61	26	53	57	49	61	51	56	48	30
D ₈	40	65	38	65	62	59	71	62	63	54	37
D ₉	55	70	54	70	72	66	73	70	71	66	51
≥ D ₉	20	5	21	5	3	9	2	5	4	9	24

Het beeld sluit opnieuw aan bij het voorgaande, hetgeen ook in overeenstemming is met de gangbare opvatting in het nijverheidsonderwijs, waar de opleiding tot electromonteur als een specialisatie binnen de metaalbewerking wordt opgevat. Een andere gangbare opvatting, dat de leerling-electromonteurs over het algemeen wat meer cerebraal zijn ingesteld dan de leerling-bankwerkers, vindt hier eveneens zijn bevestiging: het beeld sluit vooral goed aan bij dat van de instrumentmakers. Ook in dit geval vinden we hoge scores in de TH-, TN- en M-schalen tegenover een lage AV-score en een betrekkelijk lage AA- en AD-score. De L-score komt ook hier niet boven het gemiddelde uit.

10.5. *Loodgieters*

Hoewel de beroepsnaam „lood-gieter” met de inhoud van de arbeid nog

slechts zeer weinig te maken heeft, wordt zij nog steeds gehandhaafd, zowel om traditionele als om praktische redenen. De aanduiding „koper-, lood-, zinkbewerken en pijpfitten” in het programma van de l.t.s. is te langademig, benamingen als „Gawalo-monteur” en „sanitair monteur” hebben nooit populariteit gekregen. De arbeid van de loodgieter ligt in hoofdzaak op het terrein van de metaalbewerking, maar wordt in tegenstelling tot die van de bankwerker, overwegend ambachtelijk verricht. In verband met de aard van de meeste arbeidsopgaven, wordt het beroep ook nogal eens tot de bouwvakken gerekend.

Gezien deze achtergrond is het interessant, na te gaan, in hoeverre het voorkeurspatroon van de leerling-loodgieters afwijkt van b.v. dat van de bankwerkers. Dit is ook om een andere reden belangwekkend: het beroep is zeer weinig in trek. Reeds vele jaren geeft de recrutering van een voldoende aantal leerlingen grote moeilijkheden. We kunnen ons daarom afvragen, of degenen, die toch het beroep kiezen, gekenmerkt worden door een zeer geprononceerde belangstellingskarakteristiek, of dat we eerder met jongens met vrij neutrale belangstelling te maken hebben, die — wellicht door overreding — bereid zijn deze richting te kiezen. Tabel 10.4 geeft een overzicht van de bevindingen bij een groep van 18 eerstejaars en 14 tweedejaars leerlingen.

TABEL 10.4. Belangstellingskarakteristiek van leerling-loodgieters (N = 32)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	0	1	1	5	0	2	6	1	1	0	3
D ₂	0	6	5	7	1	2	7	6	3	3	8
D ₃	1	11	8	10	6	5	11	14	4	5	12
D ₄	3	14	10	14	9	10	14	17	7	8	20
D ₅	7	18	15	18	14	13	17	25	14	9	25
D ₆	9	21	20	19	18	16	22	27	15	14	27
D ₇	13	24	25	22	23	22	24	28	22	18	29
D ₈	17	26	28	24	26	26	29	29	26	23	30
D ₉	22	29	31	29	32	29	31	30	28	28	32
≥ D ₉	10	3	1	3	0	3	1	2	4	4	0

Hoewel door het betrekkelijk kleine aantal proefpersonen enige voorzichtigheid met het trekken van conclusies nodig is, zien we toch wel een aantal duidelijke tendenties. De gemiddelde TH-score is hoog, evenals bij de reeds besproken metaalbewerkers, maar de TN-score komt niet boven de doorsnee uit. De LG- en M-scores zijn aan de lage kant en met name ook lager dan bij de andere metaalbewerkers. Ook de gemiddelde AD-score is laag. De overeenstemming met het profiel van de bankwerkers in de HA- en L-schalen is wel suggestief, maar statistisch niet betrouwbaar. Dit laatste geldt ook voor het wat hoge gemiddelde op de SO-schaal.

Terwijl we bij de instrumentmakers de karakteristiek konden beschrij-

ven als een meer cerebraal-gerichte variant van de bankwerkers, krijgen we hier juist het tegendeel. De instelling is nog wat minder op de theorie en het intellectuele gericht, terwijl de toewending tot de metaalbewerking blijft bestaan. Degenen, die deze richting kiezen, kunnen dus gemiddeld niet als amorf van interesse worden beschreven. Het beroep appelleert misschien nog het meeste aan degenen, die zich tot de metaalbewerking aangetrokken voelen, maar van de theorielessen op de lagere technische school weinig gediend zijn.

Vatten we de uitkomsten betreffende de vier metaalbewerkersberoepen samen, dan blijkt er een ontwikkelingsreeks te bestaan. De karakteristieke kenmerken, zoals hoge TH-, TN- en M-score, lage AV- en AA-score komen het minst duidelijk bij de loodgieters tot uiting en vervolgens steeds meer geprononceerd bij de bankwerkers, electromonteurs en instrument-makers.

10.6. Schilders

De thans te bespreken groep werkt met een totaal ander medium en met totaal andere technieken dan de metaalbewerkers. Er bestaat integendeel eerder in vele opzichten affiniteit tussen het ambachtelijk schilderen en het kunstschilderen, een verwantschap die verder gaat dan het gebruik van hetzelfde medium. Op de lagere technische school krijgen de leerling-schilders een onderwijs, dat sterk afwijkt van dat van de meeste andere groepen, onder anderen doordat wiskunde en andere exacte vakken op het programma ontbreken. Dit heeft nogal eens tot gevolg, dat de richting wordt gekozen door zwakbegaafde leerlingen. Mede daardoor is niet bij voorbaat duidelijk, welke belangstellingskarakteristiek we moeten verwachten: overweegt de toewending tot het artistieke, of zal meer het amorf, ietwat naar het feminiene gaande patroon van de betrekkelijk zwakbegaafde groep op de voorgrond treden? De onderzochte groep bestond uit 19 eerstejaars en 19 tweedejaars leerlingen.

TABEL 10.5. Belangstellingskarakteristiek van leerling-schilders (N = 38)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	10	1	4	4	0	10	13	1	0	1	1
D ₂	19	1	12	11	3	16	15	2	3	6	5
D ₃	27	2	20	12	5	17	22	8	8	6	9
D ₄	31	2	25	17	10	23	26	10	14	12	9
D ₅	33	2	30	20	15	25	29	12	16	14	12
D ₆	34	3	32	21	21	29	33	13	22	17	18
D ₇	36	5	33	25	22	32	33	18	27	22	22
D ₈	37	8	34	28	28	35	36	21	28	28	27
D ₉	37	13	38	31	36	37	37	23	31	31	33
≥ D ₉	1	25	0	7	2	1	1	15	7	7	5

Het beeld dat deze groep vertoont, is zeer karakteristiek. Twee-derde van de onderzochte groep heeft een AV-score, die tenminste gelijk is aan het negende deciel van de totale populatie, scores die lager dan gemiddeld zijn komen vrijwel niet voor. Ook de LG-score is opvallend hoog. We mogen deze beide kenmerken wel specifiek achten voor de affiniteit tot het kunstzinnige. Daarentegen zijn de scores op de TH-, TN-, HA- en AD-schalen duidelijk aan de lage kant. Er is dus speciaal ten opzichte van de metaalbewerkers een scherpe afgrenzing. De scores op de SO- en VB-schalen zijn normaal, de groep van de schilders onderscheidt zich dus ook duidelijk van de intellectuele achterhoede met een naar het feminiene neigende karakteristiek.

10.7. *Straatmakers*

Een opleiding tot straatmaker bestaat binnen het kader van het normale programma van de lagere technische scholen niet. Op sommige scholen wordt echter wel ten behoeve van degenen, die in een leerlingstelsel tot straatmaker worden opgeleid, een cursus gegeven. Deze leerlingen zijn uiteraard iets ouder dan degenen, die een normale opleiding op een l.t.s. beginnen, omdat zij niet meer onder de leerplicht mogen vallen. Tabel 10.6 geeft een overzicht van de scores behaald door 21 leerling-sstraatmakers, waarvan 14 uit het eerste en 7 uit het tweede jaar.

TABEL 10.6. Belangstellingskarakteristiek van leerling-sstraatmakers (N = 21)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	0	3	3	1	0	2	0	1	0	1	4
D ₂	3	4	3	2	3	2	1	3	1	4	4
D ₃	4	6	6	3	6	2	2	3	3	8	4
D ₄	4	9	11	4	6	4	5	7	5	9	6
D ₅	14	16	14	7	8	5	6	10	5	9	11
D ₆	18	17	14	9	12	12	8	13	5	14	17
D ₇	18	18	16	10	12	14	11	16	8	14	20
D ₈	19	18	18	14	17	15	14	19	12	15	21
D ₉	21	19	21	17	20	17	18	21	16	19	21
≥ D ₉	0	2	0	4	1	4	3	0	5	2	0

Gezien het kleine aantal individuen hebben alleen betrekkelijk grote afwijkingen van het normale patroon significantie. In dit geval is de SO-score duidelijk hoger dan gemiddeld, de VB-score bereikt bij een t-toets op het gemiddelde ongeveer het 10% significantieniveau. De gemiddelde M-score blijkt significant laag te zijn. De overige variaties hebben uiteraard geen betekenis. In negatieve zin laat dit wel conclusies toe, omdat we bij andere betrekkelijk kleine groepen soms wel zeer geprononceerde afwijkingen van het gemiddelde hebben zien optreden. Over het geheel past het door ons gevonden beeld het beste bij de karakteristiek van de jongens, die na de zesde klasse rechtstreeks naar de praktijk gaan: ook hier

een vrij neutraal beeld, waarin alleen enkele schalen, die in het algemeen meer door meisjes worden gekozen, een relatief hoge score hebben.

10.8. Houtbewerkers

Het ligt wel voor de hand, dat tussen de beide categorieën houtbewerkers: timmerlieden en meubelmakers een aanzienlijke verwantschap in psychologische karakteristiek zal bestaan. Zij gebruiken een grotendeels identiek medium en verrichten voor een belangrijk deel dezelfde werkzaamheden. Teneinde na te gaan of er desondanks tussen de timmerlieden en meubelmakers verschillen bestaan, hebben wij de 32 timmerlieden en 27 meubelmakers uit ons materiaal afzonderlijk getabelleerd (tabel 10.7 en 10.8).

TABEL 10.7. Belangstellingskarakteristiek van leerling-timmerlieden (N = 32)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	2	1	2	7	0	3	3	0	0	2	2
D ₂	6	2	5	11	2	7	4	2	8	8	5
D ₃	8	4	9	15	8	10	10	2	11	11	8
D ₄	13	5	13	19	11	14	12	6	15	15	10
D ₅	14	7	14	25	15	15	13	10	19	18	13
D ₆	19	14	20	25	18	21	15	14	19	21	17
D ₇	25	17	27	27	21	25	18	20	20	25	18
D ₈	28	20	28	27	23	29	23	26	24	26	23
D ₉	31	24	30	31	29	31	28	28	28	31	29
≥ D ₉	1	8	2	1	3	1	4	4	4	1	3

Het beeld van de belangstelling der leerling-timmerlieden is weinig specifiek. De VB-score is aan de lage kant, terwijl er een suggestie is, dat de AV-, AD- en LG-scores wat hoger dan gemiddeld zijn. Gezien de bescheiden omvang van de steekproef en het feit, dat de afwijking van het gemiddelde niet zeer groot is, is echter de betrouwbaarheid van de constatering nog dubieus.

TABEL 10.8. Belangstellingskarakteristiek van leerling-meubelmakers (N = 27)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	2	1	4	4	2	8	5	0	3	1	3
D ₂	5	2	8	6	4	10	8	4	6	3	5
D ₃	8	2	12	12	11	13	11	7	7	6	8
D ₄	10	2	12	14	12	15	13	11	8	6	12
D ₅	12	3	14	18	13	16	15	14	11	7	15
D ₆	15	3	17	19	16	17	19	14	15	11	18
D ₇	19	4	20	19	20	18	20	17	18	15	22
D ₈	20	10	25	24	21	25	22	19	21	20	23
D ₉	23	17	27	25	26	26	25	21	23	26	24
≥ D ₉	4	10	0	2	1	1	2	6	4	1	3

Bij de meubelmakers is de belangstellingskarakteristiek meer geprononceerd dan bij de timmerlieden. Dit blijkt vooral uit de grote concentratie van hoge scores in de AV-sector. Overigens is er inderdaad veel verwantschap met de timmerlieden, opnieuw is de VB-score aan de lage kant, terwijl er weer een indicatie is dat de LG-score iets hoger dan gemiddeld ligt. Deze tendenties komen duidelijker tot uiting, wanneer de groep der houtbewerkers als geheel wordt beschouwd. De tabel van de belangstellingskarakteristiek zal hier niet worden gereproduceerd, omdat deze gemakkelijk uit de tabellen 10.7 en 10.8 kan worden samengesteld, door telkens de twee corresponderende termen uit deze tabellen op te tellen. De ligging van de medianen der schalen wordt dan: TH beneden het zesde deciel van de populatie, AV beneden het achtste deciel, TN beneden het zesde deciel, VB beneden het vierde deciel, VB beneden het vierde deciel, AA beneden het zesde deciel, HA beneden het vijfde deciel, AD beneden het zesde deciel, LG beneden het zevende deciel, SO, L en M beneden het zesde deciel. De drie karakteristieke schalen zijn dus AV hoog, LG vrij hoog, VB laag.

Ten aanzien van de reactie op de AV en LG schalen tonen de houtbewerkers duidelijke verwantschap met de schilders, alleen zijn de reacties minder extreem. Hier staat tegenover, dat de houtbewerkers geen ongevoen lage scores op de TH-, TN- en AD-schalen produceren. Binnen de groep van de houtbewerkers tonen de meubelmakers het voorkeurspatroon, met name in de reactie ten opzichte van de meest karakteristieke AV-schaal, het duidelijkst.

We kunnen dus, naast die van de metaalbewerkers, een tweede ontwikkelingsreeks constateren, die van de neutrale straatmakers en de weinig gedifferentieerde timmerlieden, via de meubelmakers naar de schilders loopt.

10.9. *Bakkers*

De door ons onderzochte groep van 87 eerstejaars leerlingen voor het bakkersbedrijf, bestaat in hoofdzaak uit leerling-banketbakkers. De overigen zullen meest in het „gemengde bedrijf” komen. Tabel 10.9 geeft een overzicht van hun scores.

We ontmoeten in de (banket)bakkers weer een groep met een zeer specifieke karakteristiek. De zeer hoge VB-score differentieert hen van alle tot dusver besproken categorieën. Evenals bij de aspirant-agrariërs geldt ook hier, dat dit belangrijkste kenmerk van het voorkeurspatroon niet alleen een aanwijzing is voor de synchrone validiteit, maar ook voor de begripsvaliditeit.

Ongeveer twee-derde van alle proefpersonen uit deze groep heeft een VB-score die op het negende deciel of hoger valt, 90% heeft een score, die niet beneden het zevende deciel van de populatie ligt. De andere hoge scores treffen we aan in de sectoren AV, LG en SO en op de L-schaal.

TABEL 10.9. Belangstellingskarakteristiek van leerling-(banket)bakkers (N = 87)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	23	1	11	0	3	3	16	1	2	1	1
D ₂	34	8	24	1	12	9	23	8	7	3	7
D ₃	48	12	34	3	23	20	33	15	16	3	16
D ₄	61	18	46	5	35	36	47	24	21	4	27
D ₅	73	29	56	6	44	42	52	36	27	6	43
D ₆	81	39	63	6	54	53	64	39	31	11	50
D ₇	85	47	71	9	66	63	69	47	43	17	62
D ₈	86	55	77	18	74	71	76	60	49	36	72
D ₉	86	67	83	35	83	80	84	77	65	61	84
≥ D ₉	1	20	4	52	4	7	3	10	22	26	3

Voor de drie eerstgenoemde kenmerken geldt, evenals voor de hoge VB-score, dat zij meer bij meisjes dan bij jongens voorkomen. Het beeld wordt dus behalve door de specifieke VB-score ook gekenmerkt door een zeer geprononceerde feminiene karakteristiek. Lage scores treffen we aan op de typisch masculiene TH- en TN-schalen en voorts op de AD-schaal.

10.10. Typografen

De meeste grafische vaklieden worden in het kader van een leerlingstelsel gevormd. Er zijn echter enkele lagere technische en lagere grafische vakscholen, waar een schoolopleiding voor zetter en drukker wordt gegeven. De leerlingen van één van deze scholen zijn in ons onderzoek betrokken. Onderzocht werden 44 jongens, waarvan 21 leerling-zetters en 23 leerling-drukkers.

De belangrijkste vraag die hier rijst is, of er een duidelijk verschil in belangstellingskarakteristiek tussen deze beide categorieën bestaat. Bij de bespreking van het werk van IRLE (paragraaf 2.3), die een aantal boekdrukkers had onderzocht, hebben wij erop gewezen, dat de door hem verkregen uitkomsten in strijd zijn met een algemeen aanvaarde mening over de geaardheid van de zetters en de drukkers. De vraag rees, of de zogenaamde boekdrukkers in werkelijkheid zetters zouden zijn geweest. Deze veronderstelling kon in dat verband niet nader worden geverifieerd, omdat gegevens over andere grafische vaklieden ontbraken. De tabellen 10.10 en 10.11, die de scoreverdeling van de zetters, respectievelijk drukkers weer geven kunnen voor de oplossing van dit probleem echter belangrijke aanwijzingen leveren.

De betrekkelijk hoge gemiddelde LG-score bij de zetters ligt geheel in de lijn der verwachting. Minder voor de hand ligt, dat ook in de AV-schaal een zeer duidelijke top wordt bereikt. Zeer uitgesproken lage scores zijn er niet, alleen de TH- en AD-scores wijken naar beneden significant van het gemiddelde af.

TABEL 10.10. Belangstellingskarakteristiek van leerling-zetters (N = 21)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	2	0	0	4	1	2	4	0	1	0	1
D ₂	2	3	4	6	5	2	6	1	1	0	2
D ₃	7	3	6	8	7	4	7	1	4	1	5
D ₄	12	4	9	10	8	7	8	4	7	3	7
D ₅	16	6	11	12	11	8	11	5	8	5	9
D ₆	19	7	15	12	13	12	13	6	12	10	10
D ₇	19	8	17	13	16	17	17	9	13	11	11
D ₈	19	8	20	15	18	17	21	12	16	13	15
D ₉	21	11	20	17	21	20	21	16	20	19	19
≥ D ₉	0	10	1	4	0	1	0	5	1	2	2

TABEL 10.11. Belangstellingskarakteristiek van leerling-drukkers (N = 23)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	2	1	0	4	0	3	5	0	1	3	1
D ₂	3	3	2	5	1	3	7	1	4	4	2
D ₃	8	3	6	10	4	5	14	2	7	7	2
D ₄	12	3	8	13	8	8	16	2	10	8	5
D ₅	14	6	11	14	9	10	17	5	10	8	8
D ₆	19	7	11	17	11	15	19	5	14	10	9
D ₇	21	8	14	19	14	16	20	9	17	11	12
D ₈	21	14	18	23	19	18	22	12	18	15	17
D ₉	21	15	19	23	22	22	23	19	18	20	18
≥ D ₉	2	8	4	0	1	1	0	4	5	3	5

Het beeld van de interesses der leerling-drukkers vertoont een verrassend grote overeenkomst met dat van de leerling-zetters. Ook hier treffen we een top in de LG- en AV-schalen aan, opnieuw zijn de AD- en TH-score lager dan gemiddeld. Vooral dit laatste is een belangrijke bevinding, omdat traditioneel aan de drukkers altijd een technische belangstelling wordt toegedacht.

Gezien de grote mate van overeenstemming in de belangstellingskarakteristieken mogen we de tabellen 10.10 en 10.11 verenigen. Evenmin als bij de houtbewerkers zal deze gecombineerde verdeling hier worden gereproduceerd, omdat zij gemakkelijk uit de beide genoemde tabellen is af te leiden. De ligging van de mediale score op de opeenvolgende schalen is: TH beneden het vierde deciel, AV op het achtste deciel, TN op de mediaan, VB beneden het vierde deciel, AA en HA beneden het zesde deciel, AD beneden het vierde deciel, LG beneden het achtste deciel, SO beneden het zesde deciel, L op het zevende deciel, M beneden het zevende deciel. In de gecombineerde verdeling overheerst dus de karakteristiek

van de drukkers, de afwijkingen zijn nu door het groter aantal proefpersonen uiteraard meer significant.

Het beeld dat we van de grafische vaklieden hebben gekregen toont in veel opzichten gelijkenis met de belangstellingskarakteristiek van de schilders, zowel in de toewending tot de AV- en LG-sectoren en de vrij hoge L-score, als in de afwending van de TH- en AD-richtingen. Alleen zien we hier, dat de TN-score gemiddeld normaal is, terwijl deze bij de schilders uitgesproken laag was. De belangstellingstest leert, dat er — tegen de verwachting in — in de onderzochte sectoren geen verschil van betekenis tussen zetters en drukkers bestaat, zodat aan de testscore geen aanwijzingen voor de differentiatie tussen deze beide richtingen kan worden ontleend.

10.11. *Leeftijdsinvloeden*

Door IRLE zijn een aantal groepen leerling-vaklieden onderzocht, die overeenkomen met de proefgroepen uit het hier gerapporteerde onderzoek, met dien verstande, dat de door IRLE onderzochte jongens een aantal jaren (gemiddeld ongeveer drie jaar) ouder zijn dan onze proefpersonen. Een vergelijking van de uitkomsten van de beide onderzoekingen zou de mogelijkheid openen, na te gaan, welke rol leeftijdsfactoren in de test spelen. In het verslag van zijn onderzoek heeft IRLE gegevens meegedeeld over de rekenkundige gemiddelden van de scores, die zijn proefgroepen op de schalen behaalden. Doordat geen gegevens over de spreiding bekend zijn, zal de vergelijking uiteraard alleen op zeer globale wijze kunnen geschieden. Wij zullen daarbij bovendien de correcties, die wij in de test hebben aangebracht, buiten beschouwing moeten laten en tenslotte ook moeten afzien van de invloed van nationale verschillen. Deze beide laatstgenoemde bezwaren behoeven niet zeer zwaar te worden geteld: de correcties zijn van ondergeschikte aard ten opzichte van het geheel en hebben bovendien alleen betrekking op items, die niet homogeen zijn met hun schaal. De structuur van de onderzochte beroepen is in Duitsland en Nederland vrijwel gelijk, ook de maatschappelijke waardering is analoog, zodat het niet a priori waarschijnlijk is, dat er in de beide culturen een groot verschil in perceptie en daardoor ook in toewending tot de beroepen zal bestaan.

Teneinde de vergelijking praktisch te kunnen doorvoeren hebben wij van alle groepen uit dit hoofdstuk, waarvan een tegenhanger in het onderzoek van IRLE te vinden was, de gemiddelde score per schaal getransformeerd in een percentielscore, die ontleend werd aan de normtabel van IRLE. In de handleiding voor zijn test ⁽⁶⁷⁾ geeft IRLE de overeenkomstige getallen voor zijn proefgroepen. In tabel 10.12 volgt een vergelijking van de percentielscores van de (rekenkundige) schaalgemiddelden voor zeven paren van beroepsopleidingen. Van ieder paar is de Duitse groep dus als

regel gemiddeld 16 jaar, terwijl de Nederlandse groep gemiddeld ongeveer 13 jaar oud is.

TABEL 10.12. Vergelijking van zeven beroepsgroepen ten aanzien van leeftijdsinvloeden

	TH	AV	TN	VB	AA	HA	AD	LG	SO	N
Schlosserlehrling	82	36	63	47	44	56	59	52	41	(19)
bankwerkersleerling	85	40	52	57	46	56	35	48	45	(103)
Feinmechanikerlehrling	85	42	74	56	45	37	37	44	29	(63)
instrumentmakersleerling	91	32	70	47	34	46	38	50	35	(51)
Elektrikerlehrling	80	40	76	52	47	53	44	47	39	(102)
electromonteurleerling	85	36	65	56	46	53	35	44	38	(75)
Zimmermannlehrling	56	61	65	54	49	55	59	45	38	(16)
timmermansleerling	70	68	46	41	63	47	45	57	52	(32)
Tischlerlehrling	74	79	43	56	63	57	32	38	32	(47)
meubelmakersleerling	73	78	36	54	54	43	40	57	52	(27)
Konditorenlehrling	49	77	40	91	62	48	43	58	42	(56)
(banket)bakkersleerling	50	63	37	92	53	52	34	55	64	(87)
Buchdruckerlehrling	35	64	60	42	43	46	61	87	37	(19)
boekdrukkersleerling	64	69	51	51	62	51	13	69	52	(23)

Bij inspectie van deze tabel blijken drie algemene tendenties. De AD- en TN-scores zijn bij de oudere leerlingen gemiddeld hoger, terwijl de SO-score gemiddeld lager ligt. Het eerste kenmerk kan gemakkelijk worden verklaard uit de algemene aversie tegen kantoorwerk bij de door ons bestudeerde leeftijdsgroep. Het is bekend, dat deze aversie met het ouder worden geleidelijk iets afneemt. De lage SO-score hangt wellicht samen met de relatief sterk introverte instelling omstreeks het midden van de puberteit. Uit Amerikaanse onderzoeken is bekend, dat de belangstelling voor de sociaal-pedagogische sector bij het naderen van de volwassen leeftijd weer toeneemt. De relatief hoge TN-score is minder gemakkelijk te verklaren. Het verschil tussen de leeftijdsgroepen is meestal niet groot, maar de richting van het verschil is consistent. Wellicht moet het worden geïnterpreteerd als een symptoom van de meer bewuste confrontatie van de oudere puber met de huidige, sterk technologisch georiënteerde wereld.

In het algemeen blijkt er een zeer grote mate van overeenstemming tussen de corresponderende groepen te bestaan, de toppen en dalen van de schaalgemiddelden liggen vrijwel steeds op dezelfde plaatsen. Alleen de groep timmerlieden vormt hier in vele opzichten een uitzondering. De karakteristiek van onze timmerlieden sluit echter goed aan bij die van onze meubelmakers, die op hun beurt weer nauwe verwantschap met de „Tischler” vertonen, terwijl voor de Duitse timmerlieden een dergelijke verwant-

schap niet bestaat. Gezien bovendien het feit, dat de door IRLE onderzochte groep slechts van geringe omvang is, menen we, dat in dit geval aan onze uitkomsten een grotere waarde mag worden toegekend.

Samenvattend kunnen we stellen, dat de leeftijdsinvloeden in de B I T zich althans in de eerste jaren na het verlaten van de lagere school voornamelijk openbaren in de AD-, TN- en de SO-schalen. De relatie tussen de overige schalen en het voorkeurspatroon van de onderzochte vakrichtingen is vrijwel constant. Dit betekent derhalve, dat de bevindingen uit dit hoofdstuk in vele gevallen mogen worden gegeneraliseerd, ook voor andere leeftijdsgroepen dan degenen, die juist de lagere school hebben verlaten.

HOOFDSTUK XI

PREDICTIEVE VALIDITEIT

11.1. *Probleemstelling*

In de beide voorgaande hoofdstukken zijn de karakteristieke belangstellingspatronen van een aantal groepen, die in het voortgezet onderwijs worden gevormd, opgespoord. Daarbij is gebleken, dat de B I T inderdaad een differentiatie tussen opleidingsgroepen mogelijk maakt. We vinden, zowel in de zesde klas bij de aspiranten voor de verschillende richtingen, als in het begin van het voortgezet onderwijs, significante verschillen tussen de diverse groepen. In vele gevallen zijn deze verschillen zo groot, dat zij ook praktische betekenis hebben.

Op grond van de zo verkregen inzichten kan men „determineertabellen” samenstellen, waaruit op eenvoudige wijze kan worden afgelezen, naar welke richtingen in het voortgezet onderwijs een geconstateerde afwijking van het gemiddelde op een bepaalde schaal eventueel verwijst. Voorbeelden van determineertabellen, gebaseerd op de gegevens die in dit boek zijn gerapporteerd, zijn als bijlagen II t.m. IV opgenomen. Vanzelfsprekend kunnen dergelijke tabellen worden uitgebreid en eventueel herzien, naarmate meer gegevens beschikbaar komen.

Gezien de resultaten van de onderzoeken, die in de hoofdstukken IX en X zijn gerapporteerd, zijn wij van oordeel, dat de B I T in het kader van de differentiatie aan het einde van de lagere school praktisch kan worden toegepast. Door groepen leerlingen te selecteren, die naar belangstellingskarakteristiek passen bij bepaalde richtingen in het voortgezet onderwijs, zou men bereiken, dat in het voortgezet onderwijs homogener groepen worden gevormd dan tot nu toe. Het is psychologisch waarschijnlijk, dat dit een positieve invloed zou hebben op de mate van bevrediging, die de opleiding de leerling geeft (¹²⁷, hoofdstuk 15).

Daarmee is echter nog niet gezegd, dat de leerlingen, die conform hun belangstellingskarakteristiek in een bepaalde opleiding worden geplaatst, daar ook meer kans van slagen zouden hebben dan anderen, die een afwijkend belangstellingspatroon vertonen. Er zijn natuurlijk in de eerste plaats andere factoren, waarvan reeds bekend is, dat zij invloed hebben op het schoolsucces. Maar zelfs wanneer wij kinderen zouden vergelijken, die naar intelligentie, werkhouding, sociaal milieu e.d. weinig van elkaar verschillen, mogen we niet a priori stellen, dat een afwijking van het „gewenste” belangstellingspatroon een slechtere prognose ten aanzien van

het schoolsucces zou geven. Het klinkt wel aannemelijk, dat een dergelijke slechtere prognose inderdaad zou gelden, maar dit dient toch eerst te worden bewezen. M.a.w. we hebben in het voorgaande wel kunnen constateren, dat de test in tal van gevallen synchrone validiteit bezit, maar of de B I T ook predictieve validiteit heeft, moet nader worden onderzocht.

Gezien de omvang der werkzaamheden die met een dergelijk onderzoek zijn gemoeid, hebben wij ons bij de bestudering van de predictieve validiteit van de B I T vooralsnog tot één opleidingsgroep beperkt: de v.h.m.o.-leerlingen. Nagegaan is, of bepaalde testcores een predictieve waarde hebben voor het succes in het eerste schooljaar. In dit geval zijn niet alleen de gebruikelijke elf schalen in het onderzoek betrokken, maar ook de „gewogen” schaal, waarvan de constructie in paragraaf 4.6 is besproken.

Als criterium voor het schoolsucces is de bevordering aan het einde van het eerste leerjaar gekozen. Erkend moet worden, dat dit criterium gebrekkig is. De bevorderingspolitiek is niet op alle scholen dezelfde. Sommige scholen selecteren b.v. aan het einde van het eerste leerjaar reeds vrij streng, terwijl andere dit eerst aan het einde van de tweede of derde klas doen. Enkele „gemakkelijke” scholen hebben zelfs de gewoonte, betrekkelijk zwakke leerlingen tot het eindexamen mee te nemen. Een leerling met één 4 en één 5 zal, bij overigens middelmatige cijfers, op de ene school wel, op de andere niet worden bevorderd. De ene school zal bovendien meer dan de andere geneigd zijn, pedagogische overwegingen te laten gelden of rekening te houden met bijzondere omstandigheden, zoals huiselijke moeilijkheden, ziekte e.d.

Verschillen tussen de scholen ontstaan ook, doordat de normen waarmee de prestaties worden beoordeeld, zeer aanzienlijk uiteen kunnen lopen. Eenzelfde prestatie zal dus op de ene school als onvoldoende, op de tweede als voldoende worden beoordeeld, waarbij de verschillen in waarderingscijfer niet zelden meer dan één punt zullen bedragen.

Een consequentie hiervan is geweest, dat wij in het materiaal, waarop het hierna te rapporteren onderzoek betrekking heeft, enerzijds een klas aantreffen, waarin van de 29 leerlingen slechts 2 niet werden bevorderd, en anderzijds een klas waarin van 14 leerlingen er 12 niet naar de tweede gingen. Een dergelijke divergentie wijst toch wel op het bestaan van zeer grote verschillen in het proces, dat op de scholen aan de bevordering ten grondslag ligt. Vanzelfsprekend wordt daardoor ook de mogelijkheid tot differentiatie, op basis van een onderzoek, dat op verschillende scholen betrekking heeft, verminderd. Onder degenen, die aan het einde van de eerste klas bevorderd worden, bevinden zich kinderen, die op een andere school beslist zouden zijn afgewezen en omgekeerd. Op grond van deze overweging zullen wij onze verwachtingen ten aanzien van de mogelijkheid, effectieve onderscheidingen aan te brengen, beslist niet hoog kunnen stellen.

Desondanks verdient de bevordering als criterium toch nog de voorkeur boven andere. De meeste andere criteria, en met name de schoolcijfers, vertonen nog veel meer tekortkomingen (¹⁴¹), terwijl het hulpmiddel, waarmee men elders nog wel gunstige ervaringen heeft gehad, de schoolvoordringentest (¹⁰) in ons land voor het middelbaar onderwijs nog niet beschikbaar is. Onder deze omstandigheden is de bevordering nog de minst slechte maatstaf, die in ieder geval het voordeel heeft, dat zij eenvoudig en op on-dubbelzinnige wijze kan worden vastgesteld.

Met het onderhavige onderzoek raken wij onvermijdelijk ook aan de gehele problematiek, die er rondom het vraagstuk van het schoolsucces bestaat. Hierop zal echter niet diep worden ingegaan, wij volstaan met het maken van enkele opmerkingen.

In de eerste plaats kan men zich afvragen, of een eventueel positief resultaat van een onderzoek, zoals hier wordt ondernomen, inderdaad iets zegt over de predictieve waarde van de test. Het zou denkbaar zijn, dat we dan te maken hebben met een samenhang, die in werkelijkheid berust op de afhankelijkheid van een derde factor, b.v. de intelligentie. Van een dergelijk verband zullen we in het volgende hoofdstuk nog een voorbeeld ontmoeten. In dit geval moet echter worden opgemerkt, dat er van een dergelijke samenhang weinig is gebleken. Met name is de correlatie tussen de interessescores en de intelligentie over de gehele lijn laag, het percentage van de variantie in de interessescores, dat op grond van de kennis van het IQ kan worden voorspeld, kan in de praktijk worden verwaarloosd. Maar ook met de variabelen, die in hoofdstuk XII nog ter sprake komen, bestaat weinig verband. Indien er al sprake is van een indirect verband, dan wordt dit dus veroorzaakt door een door ons niet onderzochte factor, die dan echter ook in het geschiktheidsonderzoek vermoedelijk niet of nauwelijks zal worden betrokken. In dat geval behouden de interessescores hun predictieve waarde.

Evenmin is er aanleiding, bij voorbaat te veronderstellen, dat de groep van de middelbare scholieren reeds thans zo homogeen naar belangstellingskarakteristiek is, dat daardoor de betekenis van de belangstelling als differentiërende factor aanzienlijk zou worden beïnvloed. Het is bekend, dat in ons onderwijsstelsel de sterke tendentie bestaat, steeds een gedeelte van de toegelaten leerlingen als onvoldoende te beoordelen, ongeacht de veranderingen die het peil van het aanbod, zeker op langere termijn, ondergaat (²⁴). Er is dus in ieder geval differentiatie en deze zal uiteraard geschieden op grond van kenmerken, waarin de leerlingengroep in aanzienlijke mate heterogeen is. Uit het vooronderzoek is reeds bekend, dat een dergelijke heterogeniteit ten aanzien van de belangstellingskarakteristiek inderdaad aanwezig is, zodat althans a priori de mogelijkheid, dat deze karakteristiek een predictieve betekenis heeft, volledig open blijft. Of deze predictieve waarde, indien werkelijk bestaande, ook behouden zou blijven als bij de toelating tot het v.h.m.o. op belangstellingstype zou

worden geselecteerd, is een geheel andere kwestie. Er is aanleiding te veronderstellen, dat in dat geval de differentiatie binnen het onderwijs voortaan volgens andere gezichtspunten zou geschieden, zonder dat het aantal zittenblijvers daardoor wezenlijk zou worden beïnvloed ⁽¹⁴⁰⁾.

Tenslotte moet nog kort worden ingegaan op een schaaltechnisch probleem. In het kader van het predictief gebruik van een testschaal kunnen we te maken krijgen met tweeërlei type relatie tussen schaalscore en criterium: er kan sprake zijn van een lineair of van een curvilineair verband. Wanneer we constateren, dat de middenwaarde van de onderzochte groep op een bepaalde schaal significant afwijkt van die van de vergelijkingsgroep, dan betekent dit, dat in het geval van lineair verband een afwijking in dezelfde richting, die groter is dan die van het gemiddelde (c.q. mediale) individu, positief moet worden gewaardeerd. Dit type relatie ontmoeten we als regel, wanneer we capaciteitscores predictief gebruiken. Indien b.v. kan worden vastgesteld, dat als middenwaarde voor de intelligentie van een bepaalde v.h.m.o.-categorie een IQ van 125 geldt, dan plegen we een IQ van 135 als gunstig te beschouwen. Bij een curvilineair verband is echter een „te veel” even ongewenst als een „te weinig”. Dit geval doet zich met name bij karaktereigenschappen niet zelden voor. Een zeer grote mate van eerzucht is b.v. even ongewenst als een zeer geringe mate, zij het ook om geheel andere redenen. Ten aanzien van de frequentieverdeling van de schaalcores bij de criteriumgroep betekent dit, dat in het geval van lineair verband, de verdeling verschoven is in de richting, waarin de middenwaarde van die van de vergelijkingsgroep (d.w.z. hier het ijkingsmateriaal) afwijkt. Bij curvilineair verband kan de middenwaarde al of niet ten opzichte van die van de vergelijkingsgroep verschoven zijn, maar in ieder geval is de spreiding anders.

De positie van de belangstellingskarakteristiek is in deze samenhang niet onmiddellijk duidelijk; het is moeilijk, een theoretisch uitgangspunt te vinden voor de beantwoording van dit vraagpunt. We menen echter, dat er langs empirische weg wel een antwoord kan worden gevonden. Bij analyse van de tabellen in de hoofdstukken IX en X blijkt, dat vrijwel steeds, wanneer de mediale positie van een groep op een of andere schaal significant van de vergelijkingsgroep afwijkt, de gehele verdeling in dezelfde richting verschoven is. Ligt de mediale positie in de onderzochte groep b.v. boven de middenwaarde van de vergelijkingsgroep, dan is ook vrijwel steeds het aantal individuen met een score, die tenminste op het negende deciel ligt, duidelijk groter dan ca. 10%. In het algemeen vinden we geen aanwijzing van een concentratie rondom een bijzondere middenwaarde, zoals te verwachten zou zijn in het geval van curvilineair verband, maar er is eerder sprake van een verschuiving in de frequentieverdeling, waarbij de mediale positie als een „minimum” moet worden beschouwd. (Van een werkelijk minimum, daargelaten of dit precies bij de mediale positie zou moeten liggen, is natuurlijk evenmin sprake als b.v. bij het IQ,

omdat langs verschillende wegen een compensatie van het „tekort” mogelijk is). Op grond van deze overweging zullen we in dit onderzoek een lineair verband tussen score en schoolsucces veronderstellen en dus de meer dan gemiddeld afwijkende score als een positieve geschiktheidsindicatie beschouwen, indien althans de afwijking naar richting overeenstemt met die van de mediale positie.

11.2. *Experimentele opzet*

Ten behoeve van het thans te rapporteren onderzoek zijn de leerlingen uit de eerste klassen van een aantal v.h.m.o.-scholen, meest lycea, te Amsterdam, Delft en Den Haag getest. In het onderzoek, dat in het eerste kwartaal van het schooljaar plaatsvond, zijn telkens de eerste klassen in hun geheel betrokken, dus met inbegrip van degenen, die reeds eenmaal hadden gedoubleerd. In eerste instantie was dit uiteraard een uitvloeisel van het feit, dat het onderzoek klassikaal plaatsvond. De aanwezigheid van deze leerlingen werd aanvankelijk echter ook een voordeel geacht, omdat daardoor in het materiaal de groep der niet-succesvollen groter zou worden. Mede op grond van de voorgaande studie werd verwacht, dat de invloed van de leeftijdsfactor niet groot zou zijn. Bij de bewerking van het materiaal is echter gebleken, dat deze oudere leerlingen toch in verschillende opzichten niet geheel vergelijkbaar zijn met de nieuwgekomenen. Weliswaar zijn de verschillen niet groot, zodat zij bij vergelijking van de v.h.m.o.-groep met andere groepen wel verwaarloosd kunnen worden, doch binnen deze categorie spelen zij toch een zodanige rol, dat de subgroepen niet tezamen mogen worden genomen.

Naast het belangstellingsonderzoek is in een betrekkelijk klein aantal gevallen ook een intelligentie-onderzoek verricht, met behulp van de L.O.III-serie. Op grond van de uitkomsten van deze test is een IQ berekend.

Vervolgens is aan het einde van het schooljaar nagegaan, hoe de vorderingen van de leerlingen waren geweest, hetgeen dus geschiedde, door te informeren of de leerlingen waren bevorderd. Een probleem ontstond ten aanzien van degenen, die tussentijds de school hadden verlaten: het bleek moeilijk, een betrouwbare uitspraak over de oorzaken van vertrek te krijgen. Ook waren er enkele leerlingen over wie op het tijdstip van navraag nog geen definitieve beoordeling kon worden gegeven, omdat een herexamen was opgelegd. Al deze twijfelgevallen zijn eenvoudigheidshalve buiten beschouwing gelaten. Uit het materiaal, dat aanvankelijk betrekking had op 501 kinderen, werden op deze wijze 25 gevallen geëlimineerd. Onder de overblijvenden bevonden zich 85 leerlingen, die voor de tweede maal in de eerste klas zaten. Van hen werden 58 kinderen wel en 27 niet bevorderd. Voor de eerste maal in de eerste klas zaten dus 391 kinderen, waarvan er 82 niet werden bevorderd, 10 een taak kregen en 299 zonder

bezwaar werden bevorderd. De beide laatstgenoemde categorieën zijn in ons onderzoek steeds tezamen genomen.

Het materiaal, dat op deze leerlingen betrekking had, is op verschillende wijzen geanalyseerd. In de eerste plaats zijn de gegevens over het gehele materiaal, alleen naar de sexen gesplitst, ten aanzien van de elf „gewone” schalen vergeleken met de bevindingen over de kinderen uit het ijkingsmateriaal, die het v.h.m.o. zouden gaan bezoeken. Dit materiaal is bovendien aangevuld met de gegevens betreffende 59 leerlingen, afkomstig uit een onderzoek van enkele tweede klassen, hetgeen in dit geval, waar het om intergroepverschillen gaat, geoorloofd was. Op deze wijze werd aan een materiaal, van aanzienlijke omvang, een controle verkregen op de gegevens, die in paragraaf 9.2 zijn gerapporteerd.

Vervolgens zijn de gegevens over degenen, die zonder dubleren de tweede klas bereikten, geplaatst tegenover de gegevens betreffende de leerlingen, die bij het onderzoek eveneens voor de eerste maal in de eerste klas zaten, maar die bij de overgang werden afgewezen. Hierbij is uiteraard weer onderscheid gemaakt naar de sexen.

Dezelfde vergelijking is ook uitgevoerd voor de subgroep van de leerlingen van enkele scholen, die reeds in de eerste klas streng selecteren.

Tenslotte zijn nog enkele speciale onderzoeken uitgevoerd. Deze hadden in de eerste plaats betrekking op de leerlingen, bij wie ook een IQ bepaald was. Getracht is, hier „matched samples” van zittenblijvers en bevorderden te construeren. Ook is nog afzonderlijk nagegaan, welke waarde in dit verband aan de „gewogen” schaal moest worden toegekend.

11.3. Een controle op de ijkingsgroep

De resultaten van alle 560 jongens en meisjes, die in het v.h.m.o.-onderzoek waren betrokken, zijn samengevat in de tabellen 11.1 en 11.2. Deze zijn rechtstreeks vergelijkbaar met de tabellen 9.1 en 9.2 over de bestemmingsgroep v.h.m.o. in het ijkingsmateriaal.

TABEL 11.1. Belangstellingskarakteristiek van meisjes, die v.h.m.o. bezoeken (N = 230)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	29	—	62	6	55	60	9	44	72	0
D ₂	—	53	8	83	16	76	95	19	71	108	3
D ₃	30	67	—	111	37	125	119	27	103	123	6
D ₄	—	77	23	121	47	135	159	50	124	134	13
D ₅	75	90	48	150	67	164	171	63	141	152	23
D ₆	123	110	72	161	82	178	192	75	151	190	31
D ₇	163	131	106	181	115	197	205	92	166	200	53
D ₈	190	155	134	195	130	213	211	125	181	212	69
D ₉	216	185	163	210	146	222	225	157	205	221	127
≥ D ₉	14	45	67	20	84	8	5	73	25	9	103

Bij vergelijking van deze tabel met de corresponderende uit hoofdstuk IX blijkt, dat de karakteristieke trekken bij de meisjes vrijwel ongewijzigd naar voren treden. De relatief hoge gemiddelden op de schalen TN, AA, LG en M zijn gebleven, evenals de relatief lage gemiddelden op de schalen VB, HA, AD en L. Aanvullend kunnen we constateren, dat er een, overigens niet zeer krachtige tendentie is tot hoge scores op de AV- en lage scores op de SO-schaal.

TABEL 11.2. Belangstellingskarakteristiek van jongens, die v.h.m.o. bezoeken (N = 330)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	55	46	5	37	64	47	47	7	42	86	1
D ₂	88	119	14	84	111	76	65	28	92	171	1
D ₃	137	157	24	124	151	111	108	48	144	226	2
D ₄	197	195	33	160	188	146	136	67	171	242	6
D ₅	225	228	42	190	216	169	155	97	192	262	25
D ₆	259	244	65	203	241	201	186	118	212	292	31
D ₇	279	268	96	238	266	228	212	143	237	299	50
D ₈	299	280	108	273	285	254	259	183	266	313	101
D ₉	318	306	166	304	309	290	295	230	293	324	167
≥ D ₉	12	24	164	26	21	40	35	100	37	6	163

Bij de jongens is het verschil tussen de twee overeenkomstige tabellen wat groter dan bij de meisjes — tegen de verwachting in, omdat tabel 9.2 betrekking had op een aanzienlijk grotere groep personen dan tabel 9.1. Toch geldt ook in dit geval, dat de hoofdkarakteristiek dezelfde is gebleven.

De frequentie van de hoge scores op de TN-schaal is nog aanzienlijk toegenomen, terwijl de hoge gemiddelden op de LG- en M-schalen gebleven zijn. Daarentegen hebben de relatief hoge HA- en AD-scores zich niet weten te handhaven, de gemiddelden liggen hier thans in de midden-zône. De lage TH-, AA- en L-scores zijn gebleven. Aanvullend blijkt nu, dat evenals bij de meisjes ook de scores op de AV- en SO-schalen van de doorsnee afwijken, ditmaal echter beide in lage richting. We herinneren er aan, dat de lage AV-score ook reeds bij jongens, die een overgangsjaar volgen, was geconstateerd (tabel 9.13).

De verschuivingen bij onderzoek van een groter materiaal komen dus bij de jongens neer op een nog meer geprononceerde toewending tot intellectuele bezigheden en een over het geheel sterkere afwending van de praktische activiteiten. Zowel bij de jongens als bij de meisjes blijken de L- en M-scores een zeer goede differentiërende werking te hebben, hetgeen te opmerkelijker is, omdat we hebben kunnen constateren, dat de statistische betrouwbaarheid van de L-score te wensen overlaat.

11.4. *Schoolsucces in het eerste jaar in het v.h.m.o.*

Volgens de bekende C.B.S. generatiestatistiek van het v.h.m.o.⁽³⁵⁾ wordt slechts ongeveer tweederde van de leerlingen van de eerste klasse direct tot de tweede klas bevorderd. Daarbij vergeleken kunnen de prestaties van onze proefpersonen, van wie ruim éénvierde de tweede niet zonder te struikelen zal kunnen bereiken, nog „bevredigend” genoemd worden. We hebben er echter reeds op gewezen, dat er ten aanzien van de bevordering zeer grote verschillen tussen de scholen bestaan en daarvan, wat de afzonderlijke klassen betreft, reeds enkele extreme voorbeelden gegeven. Laten we de leerlingen over wie (nog) geen oordeel bekend was buiten beschouwing, dan kan worden gezegd, dat van de scholen in ons materiaal het percentage niet-bevorderden, onder degenen die voor de eerste maal in de eerste klas zaten, varieerde tussen 7 en 42%. Dit is voor ons aanleiding geweest, de gegevens betreffende enkele der „strengste” scholen afzonderlijk te analyseren. Eerst zullen echter de gegevens over alle scholen worden besproken. We geven in de tabellen 11.3 en 11.4 een overzicht van de belangstellingskarakteristiek van de jongens die voor de eerste maal in de eerste klas zaten en aan het einde van het jaar wel, respectievelijk niet werden bevorderd.

TABEL 11.3. Belangstellingskarakteristiek van jongens, die de eerste klas van het v.h.m.o. vlot doorlopen (N = 162)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	24	20	2	22	31	25	13	2	22	42	0
D ₂	35	62	7	45	53	37	22	13	51	86	0
D ₃	61	80	9	68	76	53	44	24	74	114	0
D ₄	93	101	10	87	95	70	56	34	87	125	1
D ₅	107	120	14	95	105	81	64	53	97	132	1
D ₆	122	130	23	105	114	96	81	61	110	146	10
D ₇	140	138	36	124	128	113	95	74	121	146	12
D ₈	146	144	42	136	140	127	126	89	135	155	22
D ₉	156	154	74	153	151	146	144	112	147	160	81
≥ D ₉	6	8	88	9	11	16	18	50	15	2	81

Bij vergelijking van de beide tabellen blijken er inderdaad verschillen tussen deze beide groepen jongens te bestaan. Bij vergelijking op de mediale waarden van tabel 11.3. is dit verschil significant op de TN-schaal en benadert het de eerste significantiedrempel ($P \text{ ca } .10$) op een aantal andere schalen (L, M, VB, AA). Belangrijker is misschien nog de constatering, dat de jongens die direct worden bevorderd, over de gehele lijn het karakteristieke beeld van het v.h.m.o. meer geprononceerd vertonen, dan degenen die aan het einde van het eerste jaar niet worden bevorderd. Degenen, die in de eerste klas niet blijven zitten hebben een hogere me-

TABEL 11.4. Belangstellingskarakteristiek van jongens, die aan het einde van het eerste jaar in de eerste klas van het v.h.m.o. niet worden bevorderd (N = 49)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	8	6	2	1	8	7	11	4	6	13	1
D ₂	14	16	4	7	13	14	13	8	13	20	1
D ₃	23	22	6	13	17	18	19	12	21	29	1
D ₄	30	29	8	18	23	22	22	14	25	30	1
D ₅	31	35	9	24	27	26	26	14	27	34	6
D ₆	39	37	15	25	31	33	29	17	29	37	10
D ₇	43	42	18	33	35	37	30	22	32	40	13
D ₈	46	44	20	40	37	42	38	27	38	44	19
D ₉	48	46	32	44	43	45	44	37	41	46	32
≥ D ₉	1	3	17	5	6	4	5	12	8	3	17

diale score op de TN-, AD- en M-score en een lagere VB-, AA- en L-score. Opvallend is, dat vooral bij de groep van de succesvollen het kenmerk van een hoge waarde voor de mediale AD-score voorkomt, dat we ook bij de ijkingsgroep hadden aangetroffen, maar dat bij de contrôlegroep in de vorige paragraaf niet meer aanwezig scheen te zijn. Blijkbaar heeft de invloed van de oudere leerlingen in het materiaal van tabel 11.3 bewerkstelligd, dat dit kenmerk verdween. Dit zou er op kunnen wijzen, dat we hier met een sterk leeftijdsgebonden kenmerk van de v.h.m.o.-groep te maken hebben.

TABEL 11.5. Belangstellingskarakteristiek van meisjes, die de eerste klas van het v.h.m.o. vlot doorlopen (N = 147)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	18	—	39	6	37	37	7	24	44	0
D ₂	—	35	5	54	12	48	60	14	45	68	3
D ₃	16	44	—	71	22	83	78	18	66	78	4
D ₄	—	49	16	77	27	90	108	29	81	85	7
D ₅	49	59	34	94	37	114	116	35	90	98	13
D ₆	80	70	49	105	47	121	125	45	97	120	19
D ₇	103	85	71	118	67	131	132	52	110	127	31
D ₈	122	103	84	126	78	138	135	77	118	135	41
D ₉	136	121	100	139	89	142	145	98	133	142	78
≥ D ₉	11	26	47	8	58	5	2	49	14	5	69

De verschillen tussen de beide groepen zijn bij de meisjes wat minder geprononceerd dan bij de jongens. Weliswaar zijn er ook hier argumenten aan te voeren ten gunste van de stelling, dat degenen, die direct succes boeken het karakteristieke beeld meer uitgesproken vertonen, dan de

TABEL 11.6. Belangstellingskarakteristiek van meisjes, die aan het einde van het eerste jaar in de eerste klas van het v.h.m.o. niet werden bevorderd (N = 33).

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	0	—	5	1	7	6	1	8	8	0
D ₂	—	6	1	7	3	8	11	3	11	14	0
D ₃	8	9	—	13	6	10	12	6	17	15	2
D ₄	—	11	3	16	9	10	17	10	20	15	4
D ₅	11	12	5	20	13	13	18	13	23	17	4
D ₆	18	16	7	21	16	16	25	13	25	25	5
D ₇	26	17	15	25	22	20	28	16	26	28	11
D ₈	28	19	21	29	23	25	30	20	27	31	13
D ₉	32	23	26	32	24	30	32	24	30	33	21
≥ D ₉	1	10	7	1	9	3	1	9	3	0	12

zittenblijvers. De VB-, HA-, AD- en L-scores zijn bij de succesvollen gemiddeld wat lager. De AA-score is wat hoger, terwijl op de TN-, LG- en M-schalen, waar de ligging van de medianen bij de beide groepen niet verschilt, het percentage zeer hoge scores bij de succesvollen duidelijk groter is. Hier staat echter tegenover, dat op de AV- en SO-schalen de zittenblijvers het karakteristieke beeld wat duidelijker vertonen. Past men, uitgaande van de hypothese, dat de succesvolle groep het karakteristieke beeld meer geprononceerd zal vertonen, op de gevonden verschillen een tekentoets toe, dan ligt het gevonden verschil — bij eenzijdige toetsing — op de grens van 5% significantie. Maar al zijn er dus aanwijzingen voor het bestaan van een verschil in het belangstellingspatroon van de beide groepen, toch is de onderlinge overeenstemming hier duidelijk groter dan bij de jongens.

11.5. De invloed van strenge selectie

We hebben er reeds op gewezen, dat de bevorderingspolitiek op de scholen, die in het onderzoek waren betrokken, grote verschillen moet vertonen, gezien de zeer uiteenlopende percentages bevorderden. Het ligt voor de hand hier de vraag te stellen, of op de „strenge” scholen het karakteristieke beeld van de succesvollen in versterkte mate aanwezig blijkt te zijn. Als er inderdaad een relatie tussen belangstellingspatroon en schoolsucces bestaat, dan moet worden verwacht, dat dit op de „strenge” scholen het eerst aan het licht zal komen. Op de scholen, die in de eerste klas geen grote barrières opwerpen, zal ook een leerling, die naar type niet past, nog een aanzienlijke kans hebben, te worden bevorderd. Een dergelijke leerling zal dan wellicht in een volgende klas moeilijkheden krijgen, maar aangezien ons onderzoek zich slechts over één jaar uitstrekte, zal dat hier nog niet blijken.

We hebben daarom het materiaal, dat betrekking had op de drie strengste scholen uit ons materiaal, opnieuw geanalyseerd. Ten aanzien van de zittenblijvers leverde dit uiteraard geen nieuwe gezichtspunten op, omdat ook degenen, waarop de tabellen 11.4 en 11.6 betrekking hadden, in zeer grote meerderheid van deze scholen afkomstig waren. Omdat de tabellen, die op de doubleergevallen van de strenge scholen betrekking hadden, niet opvallend van de beide hiervoor genoemde verschilden, zijn zij hier niet gereproduceerd. Wij vermelden hier alleen, in de tabellen 11.7 en 11.8, de frequentieverdelingen van de jongens, respectievelijk meisjes, die op deze strenge scholen aan het einde van het eerste schooljaar wel werden bevorderd.

TABEL 11.7. Belangstellingskarakteristiek van jongens, die op een v.h.m.o.-school met strenge selectie de eerste klas vlot doorlopen (N = 83)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	10	12	3	9	20	11	8	2	7	28	0
D ₂	15	36	4	23	31	15	13	7	24	52	0
D ₃	31	45	5	37	45	25	21	12	36	65	0
D ₄	49	59	7	45	52	36	25	17	39	68	0
D ₅	58	66	9	50	57	42	31	25	43	71	1
D ₆	67	72	13	54	61	54	39	30	49	76	4
D ₇	74	72	16	63	67	65	44	35	54	76	7
D ₈	78	75	17	72	73	68	62	42	61	80	12
D ₉	82	79	33	79	80	76	70	54	71	82	37
≥ D ₉	1	4	50	4	3	7	13	29	12	1	46

Inderdaad blijken op een aantal schalen de karakteristieke trekken, bij vergelijking met tabel 11.2 duidelijker op te treden. Het percentage zeer hoge scores op de TN- en M-schalen is aanzienlijk hoger geworden, het percentage lage scores op de L-schaal eveneens groter. Verder zien we een toenemende tendentie tot afwijking in de verwachte richting op de AV- en AD-schalen. Alleen de tendentie tot lage scores op de SO-schaal heeft zich niet gehandhaafd, hier vertoont de verdeling in tegendeel weer een gemiddeld beeld.

We kunnen tabel 11.7 ook nog vergelijken met tabel 11.4. Dan kan worden geconstateerd, dat significant meer jongens, die een eerste strenge selectie weten te doorstaan, een percentiel score van 80 of meer op de TN- en M-schalen en een score van minder dan 20 op de L-schaal behalen. Een lage AV- en AA-score en een relatief hoge score op de AD- en LG-schalen behoren eveneens tot de veel voorkomende kenmerken. De TH- en VB-score zijn zowel bij de succesvollen als bij de doubleerders gemiddeld wat beneden de doorsnee, maar zij schijnen geen betekenis voor de differentiatie tussen deze beide groepen te hebben.

TABEL 11.8. Belangstellingskarakteristiek van meisjes, die op een v.h.m.o.-school met strenge selectie de eerste klas vlot doorlopen (N = 71)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	9	—	23	2	36	15	1	16	25	0
D ₂	—	16	3	28	7	48	25	5	24	38	1
D ₃	6	21	—	37	12	57	35	6	31	42	1
D ₄	—	24	5	39	14	63	50	16	36	46	3
D ₅	26	30	11	48	22	65	53	22	41	54	4
D ₆	39	35	17	51	28	67	55	27	45	61	5
D ₇	50	40	27	53	36	68	60	30	50	64	10
D ₈	59	48	34	61	42	68	63	40	56	68	17
D ₉	67	57	44	67	47	70	69	49	63	68	32
≥ D ₉	4	14	27	4	24	1	2	22	8	3	39

Ook bij de meisjes komen nu een aantal karakteristieke trekken duidelijk tot uiting. De gemiddelde scores op de TN- en M-schalen liggen hoger, terwijl de HA-score gemiddeld opvallend laag ligt. Verder liggen ook de VB- en L-scores lager dan in tabel 11.5. Tegen de verwachting in geldt voor de AA-score niet een analoge regel, deze ligt ditmaal dicht bij de doorsnee dan in het grotere materiaal het geval was.

Bij vergelijking van bovenstaande tabel met tabel 11.6 blijkt het voornaamste differentiërende criterium de HA-score te zijn. Slechts ca. 10% van de succesvollen heeft een HA-score van 40 of hoger, terwijl 70% van de doubleergevallen een dergelijke percentielscore heeft. De overige verschillen in de tabel zijn niet significant, wel kan men echter zeggen, dat er bepaalde consistente tendenties aanwezig zijn: een TN- en M-score, die in de meeste gevallen tenminste 80 bedraagt, een L-score en een VB-score lager dan 40. Voorts is ook de AD-score als regel laag. Het valt op, dat evenals bij de jongens de SO-score bij de succesvolle kandidaten weliswaar beneden de doorsnee blijft, maar daarvan toch gemiddeld minder afwijkt dan bij de doubleergevallen.

Gezien deze bevindingen is de conclusie gerechtvaardigd, dat inderdaad zelfs binnen een geselecteerde groep, als die van de v.h.m.o.-candidaten, de interesseskarakteristiek predictieve waarde heeft, zij het dat deze bescheiden is. Het effect is het duidelijkst bij de scholen, waar reeds in de eerste klas een strenge selectie plaatsvindt.

Afgezien van het speciale geval van de HA-scores is de differentiërende werking van de belangstelling bij jongens duidelijker aan te tonen dan bij meisjes. Hiervoor kunnen verschillende verklaringshypothesen worden opgeworpen. Het is denkbaar, dat in een aantal gevallen bij de beoordeling van de schoolprestaties der meisjes andere, eventueel minder zakelijke, maatstaven worden aangelegd, dan bij de beoordeling van de jongens. Het is echter ook mogelijk, dat de belangstelling als motiverende factor bij

meisjes een andere rol speelt dan bij jongens, terwijl het tenslotte ook niet is uitgesloten, dat de test te zeer aan specifiek manlijke interessen appelleert, om bij meisjes een scherp differentiërende werking te kunnen hebben. Een definitieve uitspraak over de waarde van deze verklarings-hypothesen kan binnen het bestek van deze studie niet worden gegeven.

11.6. *De waarde van een „gewogen” schaal*

In een aantal gevallen bestond de gelegenheid, bij de leerlingen tevens een intelligentie-onderzoek te verrichten. Een dergelijk onderzoek kost, bij gebruik van een klassikale test, enkele uren, zodat met ons totale onderzoek dan tenminste een halve schooldag gemoeid was. Het is begrijpelijk, dat dit niet op alle scholen in het kader van het leerprogramma kon worden toegestaan, zodat slechts bij een honderdtal leerlingen een intelligentiebepaling mogelijk was. Uit deze leerlingen zijn paren gevormd, telkens van kinderen van gelijke sexe, die voor de eerste maal in de eerste klas zaten en van nagenoeg gelijke intelligentie waren, maar waarvan de ene partner aan het einde van het eerste leerjaar wél, de andere niet werd bevorderd. Een verdergaande gelijkmaking, b.v. door ook de leeftijd — die echter als regel niet sterk uiteenloopt — en het sociaal milieu constant te houden, bleek bij de beperkte omvang van het materiaal niet mogelijk. Het verschil in IQ binnen een paar bedraagt ten hoogste 3 punten.

Het bleek mogelijk, 16 paren te vormen, 10 van meisjes en 6 van jongens, die aan de genoemde voorwaarden voldeden. Dit aantal is uiteraard zeer gering, zodat de kans groot is, dat de werking van differentiërende factoren door toevallige invloeden wordt overdekt.

Van de 16 onderzochte paren hadden 4 een IQ van 130 of hoger, 4 een IQ tussen 120 en 129, 5 een IQ tussen 115 en 119 en 3 een IQ tussen 110 en 114. De beoordeling geschiedde, door het teken en eventueel de grootte van het verschil tussen de paren na te gaan. Beperkt men zich tot het teken van het verschil, dan kunnen in een aantal gevallen de jongens en de meisjes gecombineerd worden beoordeeld, namelijk wanneer mag worden verondersteld, dat de positief differentiërende factoren bij de beide sexen in dezelfde richting werkzaam zijn.

De hypothesen, dat bij meisjes én jongens de TN-, en M-scores bij de succesvollen hoger zijn, konden niet op significante wijze worden bevestigd, evenmin als de hypothesen, dat bij beide sexen de VB- en L-scores bij de succesvollen lager liggen.

Wel werd bevestigd, dat bij meisjes de HA-score bij de succesvollen lager ligt, het verschil was volgens de tekentoets op 5%, volgens de rangtekentoets zelfs op 2% significant. Positieve resultaten had ook het onderzoek van de hypothese, dat bij beide sexen de score op de „gewogen” MO-schaal, d.w.z. de schaal, waarin de keuzen op de hoekpunten het beoordelingscriterium leveren, bij de succesvollen hoger ligt. Bij de meisjes lagen 8 van de 10 verschillen in de verwachte richting (volgens de rang-

tekentoets bleek dit significant op 1%), bij de jongens 5 van de 6 verschillen. Doordat de frequentieverdeling van de scoreverschillen op de gewogen schaal nagenoeg geen sexeverschillen vertoont, kon op de gecombineerde sexen ditmaal de rangtekentoets worden toegepast, het gevonden verschil bleek significant op de 1%-drempel.

Dit resultaat was aanleiding, te onderzoeken, of ook bij een groter materiaal, analoog aan dat wat in de vorige paragrafen de gegevens voor de analyse heeft geleverd, de differentiërende werking van de gewogen schaal kon worden aangetoond.

Daartoe zijn de scores op de M-schaal en op de gewogen MO-schaal voor een aantal groepen eerste klassers vergeleken. De vergelijking geschiedde, door een t-toets op het verschil tussen de gemiddelden op beide schalen toe te passen.

In dit geval zijn, behalve leerlingen, die voor de eerste maal in de eerste klas zaten en aan het einde van het jaar wél (groep I) of niet (groep II) werden bevorderd, ook leerlingen, die voor de tweede maal in de eerste klas zaten, in het onderzoek betrokken. Ook in hun geval werd onderscheid gemaakt tussen degenen, die aan het einde van het jaar wél (groep III) of niet (groep IV) werden bevorderd. Bij de meisjes omvatte deze vierde subgroep echter zo weinig individuen, dat zij buiten beschouwing moest worden gelaten. Tabel 11.9 geeft een overzicht van de aantallen individuen in de verschillende subgroepen en de gemiddelden en standaarddeviaties van de ruwe scores op de beide schalen.

TABEL 11.9. Scores van groepen leerlingen uit eerste klassen van het v.h.m.o. op twee niveauschalen

	N	gewogen MO-schaal gemidd. s.d.		M-schaal gemidd. s.d.	
<i>Jongens</i>					
Groep I	50	25.2	24.3	31.7	10.8
Groep II	49	16.8	23.1	26.7	8.7
Groep III	43	25.1	22.8	31.1	8.7
Groep IV	20	20.5	26.4	28.4	13.1
<i>Meisjes</i>					
Groep I	48	16.3	14.9	26.5	8.2
Groep II	33	13.8	12.2	24.2	7.8
Groep III	15	15.7	18.5	24.8	8.5

Opmerking: Op de gewogen MO-schaal zijn negatieve scores mogelijk. Daardoor kan de standaarddeviatie eventueel groter zijn dan het gemiddelde, zonder dat de verdeling sterk van het normale afwijkt.

Bij meisjes blijken de verschillen tussen de groepen zo klein te zijn, dat zij niet significant zijn. Bij de jongens heeft op de M-schaal de toetsingsgrootheid van het verschil tussen de eerste en tweede groep een waarde

van 2.54 ($P < .02$) en van het verschil tussen de derde en tweede groep een waarde van 2.42 (eveneens $P < .02$). Op de gewogen MO-schaal bedraagt de waarde van de toetsingsgrootte voor de beide zelfde paren respectievelijk 1.76 en 1.73 ($.10 < P < .05$).

In dit geval overtreft de M-schaal dus, in differentiërende werking, de „gewogen” MO-schaal. Voorlopig zien wij daarom geen aanleiding, het gebruik van de omslachtig te hanteren gewogen schaal voor de praktijk aan te bevelen. Getracht zal echter worden, een groter materiaal te verzamelen, teneinde na te gaan of, op langere termijn, toch de gewogen schaal merites heeft, die bij dit eerste onderzoek niet voldoende zijn gebleken.

Een interessant aspect van tabel 11.9 vormen nog de prestaties van de leerlingen, die voor de tweede maal in de eerste klas zitten. Zowel bij jongens als bij meisjes heeft groep II (d.w.z. degenen die voor de eerste maal in de eerste klas zitten en aan het eind van het jaar niet worden bevorderd) lagere gemiddelde scores dan deze leerlingen. Hoewel slechts in één geval van een significant verschil kan worden gesproken, is toch het feit, dat we met een consistent verschijnsel te maken hebben, opmerkelijk. Uit groep II komen immers, indien we de kinderen uitschakelen, die na de eerste klasse de school verlaten, degenen voort, die volgend jaar tot de groepen III en IV zouden moeten worden gerekend. Vooral opvallend is, dat ook groep IV hogere scores heeft dan groep II. Dit zou, indien we niet met een toevallige eigenaardigheid van het materiaal te maken hebben en het verschil ook niet voortkomt uit het aandeel van de leerlingen, die na de eerste klas de school verlaten (een veronderstelling, die a priori niet erg waarschijnlijk is) er op wijzen, dat een jaar verblijf op de school een vormende werking op de belangstelling van de leerlingen heeft. Daardoor komt zelfs bij degenen, die het onderwijs niet met succes volgen, het karakteristieke belangstellingspatroon sterker tot ontwikkeling. Dit effect heeft weliswaar, gemeten in termen van het verschil met het doorsnee-kind van dezelfde leeftijd, weinig te betekenen, maar is toch voldoende duidelijk, om bij vergelijking binnen het v.h.m.o. tot uiting te komen. Ook deze veronderstelling verdient toetsing op langere termijn en aan een groter materiaal.

HOOFDSTUK XII

PERSOONSKENMERKEN EN B I T-SCORES

12.1. *Vraagstelling*

Bij het ijkingsonderzoek zijn over alle proefpersonen een aantal personalia verzameld: sexe, geboortejaar, woonplaats, beroepsniveau vader, onderwijsrichting („zuil”), bestemming in het voortgezet onderwijs. Van deze gegevens hebben de eerste en de laatste in de voorgaande hoofdstukken een belangrijke rol gespeeld. Zij zijn uiteraard ook degene, waarvan in het kader van een onderzoek met het oog op voorlichting bij differentiatie, de meeste aandacht zal worden geschonken.

Toch is het wel belangwekkend na te gaan, of er ten aanzien van de andere genoemde variabelen ook bepaalde verschillen in de gemiddelde wijze van beantwoording der B I T-items kunnen worden opgemerkt. In het algemeen zal de praktische betekenis van de op deze wijze verzamelde wetenschap niet zeer groot zijn, maar voor de verdere ontwikkeling van de belangstellingspsychologie kan een dergelijk onderzoek toch nuttig zijn.

Vanzelfsprekend wijst de enkele constatering van een correlatie tussen één der genoemde variabelen en een bepaalde schaalscore niet aanstonds op het bestaan van een causale relatie. Doorgaans zal het niet mogelijk zijn, op grond van de beschikbare gegevens met enige stelligheid een uitspraak te doen over de ware aard van de samenhang. Een nader onderzoek zou hier nodig zijn. Teneinde te onderzoeken of er verband bestaat tussen een bepaald persoonsmerk en een testcategorie, is steeds gebruik gemaakt van een vergelijking met het basismateriaal, dat afkomstig is uit het ijkingsonderzoek. De gegevens, die op de beide sexen betrekking hebben, zijn daarbij afzonderlijk behandeld. Ter karakteristiek van de eventueel bestaande relatie is voornamelijk de *phi-coëfficiënt* op de medianen berekend (vergelijk paragraaf 8.3); eventueel is echter ook de concrete frequentieverdeling, geprojecteerd op het decielenrooster van de basisgroep, onderzocht.

12.2. *Leeftijdsverschillen in het ijkingsmateriaal*

In paragraaf 10.11 is geconstateerd, dat bij vergelijking van overeenkomstige opleidingsgroepen van verschillende leeftijd slechts weinig verschillen in de B I T-score aan het licht treden. Vergelijking van ca. 13-jarige jongens met gemiddeld 16-jarigen wees alleen op het stijgen van de AD- en TN- en een dalen van de SO-score in dit leeftijdsbestek.

We vinden ook in ons ijkingsmateriaal kinderen van verschillende leeftijd, die allen leerling van een zesde klasse van een lagere school zijn. De oudere leerlingen onderscheiden zich echter van de jongeren, doordat zij vrijwel steeds één of meermalen hebben gedoubleerd. We mogen daarom wel bij voorbaat veronderstellen, dat de oudere leerlingen uit het ijkingsmateriaal niet in alle opzichten vergelijkbaar zijn met de jongere. Algemeen bekend is, dat het intelligentieniveau bij de zittenblijvers gemiddeld lager ligt. Bij een onderzoek in Noord-Brabant ⁽⁸⁾ bleek, dat het IQ bepaald met behulp van een verbale intelligentietest (L.O.IV-serie) bij degenen, die steeds waren bevorderd gemiddeld 104.35, bij degenen die eenmaal hadden gedoubleerd 98.9 en bij degenen, die twee of meermalen hadden gedoubleerd 95.5 bedroeg. Bij een intelligentiebepaling met behulp van een non-verbale test („Progressive Matrices”) waren de verschillen iets kleiner, de gemiddelden bedroegen nu achtereenvolgens 105.0, 100.5 en 99.15. Hoewel de correlatie tussen de schaalscores en het verbale IQ niet groot is (vergelijk paragraaf 4.7), mag deze bij een detailstudie niet worden verwaarloosd.

Nagegaan is nu, of er verschillen in de voorkeur bij de diverse leeftijdsgroepen in de zesde klasse voorkomen. In ons materiaal, dat uit 487 meisjes en 531 jongens bestaat, bevonden zich 244 kinderen, die in het lopende schooljaar tenminste de leeftijd van 13 jaar bereikten. Dit waren 106 meisjes en 138 jongens, zodat de percentverhouding van de beide sexen onder deze ouderen voldoende overeenkomt met die in het gehele materiaal ($\chi^2 = 2.47$, $P > .10$).

Een vergelijking, voor de beide sexen afzonderlijk, leverde weinig sprekende verschillen tussen de ouderen en de jongeren op. Als enige kan worden gereleveerd, dat de voorkeur van de oudere meisjes voor de items uit de voedselbereidingssector significant groter was ($\phi = +.14$). Aangezien de sexeverhoudingen in de beide groepen gelijk zijn, is vervolgens nagegaan of bij het grotere gecombineerde materiaal nog duidelijke verschillen aan de dag zouden treden. Dit bleek inderdaad het geval te zijn. De oudere leerlingen hadden een significant lagere score op de LG- en de M-schaal (in beide gevallen $\phi = -.13$) en een significant hogere score op de L-schaal ($\phi = +.12$). Bovendien was ook de HA-score wat hoger ($\phi = +.06$, $P < .05$). Al de genoemde verschillen zijn in de richting, die ook op grond van het intelligentieverschil moest worden verwacht, zodat in eerste instantie de veronderstelling gerechtvaardigd schijnt, dat het — overigens betrekkelijk geringe — onderscheid tussen de leeftijdsgroepen in de zesde klasse niet aan de leeftijdsfactor, maar aan het verschil in begaafdheidsniveau moet worden toegeschreven.

12.3. *Woongebied*

Bij het onderzoek naar de samenhang tussen de aard van woongemeente en de voorkeur, traden slechts in twee gevallen verschillen aan het licht:

enerzijds bij de kinderen, die in de zeer grote steden (meer dan 500.000 inwoners) leven, anderzijds bij degenen die op het platteland of in een „overgangsgebied” wonen. In beide gevallen heeft de vergelijking plaats gehad tussen het betreffende woongebied en het totale overige materiaal.

Bij de meisjes uit de zeer grote steden waren in twee sectoren de scores duidelijk lager dan bij de overige meisjes: HA ($\phi = -.23$) en AD ($\phi = -.15$). Bij de jongens uit deze steden werden eveneens in twee sectoren verschillen gevonden, ditmaal betrof het hoger liggende scores op de M-schaal ($\phi = +.11$) en de VB-schaal ($\phi = +.10$).

Bij de plattelandsmeisjes was de AD-score duidelijk hoger dan gemiddeld ($\phi = +.27$), terwijl op twee andere schalen, SO ($\phi = -.16$) en AA ($\phi = -.16$) de gemiddelde score lager lag. Bij de jongens uit de plattelandsgebieden lag de TN-score lager ($\phi = -.18$), evenals de M-score ($\phi = -.13$), terwijl daarentegen de AA-score gemiddeld hoger lag ($\phi = +.11$).

Een groot deel van de genoemde verschillen correspondeert met de karakteristiek van de v.h.m.o.-groepen ten opzichte van het vergelijkingsmateriaal. Of het gevonden verschijnsel echter zou moeten worden verklaard uit het inderdaad aanzienlijk grotere percentage van de grotestadsjeugd, dat naar het v.h.m.o. gaat, blijft een open vraag. Het kan ook zijn, dat de jeugd in de grote steden over het geheel een meer „theoretische” instelling heeft en het is zelfs niet uitgesloten, dat de gevonden verschillen voortkomen uit de verschillen, die er in allerlei andere opzichten tussen de grote stad en het platteland bestaan.

12.4. Beroepsniveau vader

Bij de niveau-indeling van de beroepen der vaders van de onderzochte kinderen is het schema van TOBI en LUYCKX⁽¹³⁰⁾ gevolgd. Hierin worden negen niveau's onderscheiden van 1 (zeer eenvoudige arbeid op handlangersniveau) tot 9 (hogere beroepen, die in het algemeen een academische scholing veronderstellen). Op grond van deze indeling kunnen de kinderen onderling worden vergeleken, waarbij zowel afzonderlijke niveau's als combinaties van opeenvolgende niveau's tegenover de rest kunnen worden geplaatst. Het aantal vergelijkingsmogelijkheden wordt op deze wijze zeer groot, vooral als men in aanmerking neemt, dat ieder vergelijkingspaar ten aanzien van de verschillen op elf schalen kan worden getoetst. Er zijn 30 verschillende mogelijkheden van afzonderlijke of combinaties van opeenvolgende niveau's, die telkens op 11 schalen ten opzichte van de rest kunnen worden getoetst, zodat er in principe 330 toetsingsmogelijkheden zijn.

Niet alle 30 niveaugroepen zijn echter even zinvol; het lijkt b.v. weinig zin te hebben om de niveau's 7 + 8 te vergelijken met de rest (d.w.z. de groepen 1 t/m 6 en 9). Maar zelfs als men enige reductie toepast, blijven er nog vele toetsingsmogelijkheden en dus ook een niet geringe kans op het

optreden van extreme toevallige variaties. DE GROOT (⁵⁴) heeft er op gewezen, dat in een dergelijke situatie het gevaar dreigt, dat men de toevallige eigenaardigheden van het materiaal gaat uitbuiten. Men kan dus in dit geval alleen maar waarde hechten aan verschillen met een zeer kleine toevalskans.

Dergelijke verschillen zijn inderdaad gevonden, zij het ook slechts in vrij klein aantal. Opvallend is wel, dat deze verschillen alleen bij de jongens optreden. Bij de meisjes werd geen enkel zeer significant verschil aangetroffen. Bij combinatie van de jongens en meisjes (in dit geval weer geoorloofd, omdat de sexeverdeling op de verschillende niveau's slechts weinig uiteenloopt) werden nog wel enkele verschillen geconstateerd.

In concreto blijkt, dat de jongens, wier vaders op de drie laagste niveau's zijn ingedeeld, significant lage scores op de TN- ($\phi = -.14$) en M-schaal ($\phi = -.18$) boeken. In de middenlaag, niveau's 3 t/m 5 is enige voorkeur voor de VB-schaal ($\phi = +.12$) aanwezig. Op de drie hoogste niveau's vinden we een afwijking van de L-schaal ($\phi = -.19$), terwijl op de twee hoogste niveau's de voorkeur voor de TN-schaal een significant waarde bereikt ($\phi = +.13$).

De genoemde verschillen zullen wel in de eerste plaats een uitvloeisel zijn van de reeds lang bekende begaafdheidsverschillen (^{45, 61}), die parallel gaan aan de verschillen in beroepsniveau bij de vaders. We kunnen overigens wel concluderen, dat de invloed die hiervan op de belangstellingskarakteristiek uitgaat, praktisch slechts zeer weinig te betekenen heeft.

Bij combinatie van de beide sexen blijkt, dat op de drie laagste niveau's een significant geringer voorkeur voor de LG-schaal bestaat ($\phi = -.11$), terwijl op de vier hoogste niveau's de HA-schaal significant minder vaak wordt gekozen ($\phi = -.11$). Ook deze verschillen zullen wel vooral met begaafdheidsverschillen in verband moeten worden gebracht.

12.5. *Beroepssector vader*

Ten aanzien van de sector waarin de vader werkzaam was of is, hebben we onderscheid gemaakt tussen landbouw, industrie, handel, administratie, diensten, gezag, verkeer en vrije beroepen en onderwijs. Sommige van deze subgroepen hebben een kleine omvang, in het bijzonder wanneer ook nog een splitsing naar sexen wordt doorgevoerd. Daarom zijn in een aantal gevallen enkele kleinere groepen gecombineerd in het onderzoek betrokken. Daardoor dreigt opnieuw het gevaar, dat ook in de voorgaande paragraaf werd gesignaleerd: de toevalskans op het optreden van „significante” afwijkingen kan niet meer worden verwaarloosd. Opnieuw geldt dus, dat alleen verschillen met een zeer geringe toevalskans aanvaardbaar zijn. Bovendien kunnen we, evenals in het voorgaande geval, onderzoeken, of deze verschillen ook inzichtelijk te aanvaarden zijn.

Veel vaker dan bij de niveau-indeling werden thans verschillen gevonden. Bij de meisjes, waarvan de vader in de landbouw werkzaam is, vinden

we een duidelijke voorkeur voor de AD- en TH-schalen (ϕ resp. $+ .21$ en $+ .12$) naast een afwending van de SO-items ($\phi = - .12$). Het beeld is dus verschoven in masculiene richting. Bij de dochters van industrie-arbeiders vinden we de tegengestelde tendentie: voorkeur voor de SO-schaal ($\phi = + .10$) en afwending van de AD-schaal ($\phi = - .14$). Bij de meisjes, waarvan de vader een vrij beroep uitoefent, of in het onderwijs werkzaam is, vinden we een meer dan gemiddelde belangstelling voor de AA-schaal ($\phi = + .13$) en een indicatie (de significantiedrempel van 5% wordt wel, die van 1% niet bereikt), dat de AD-schaal weinig in tel is ($\phi = - .09$). Uitgaande van de gedachte, dat een aanzienlijk deel van de betreffende meisjes wel naar het v.h.m.o. zal gaan, zijn deze beide bevindingen wel begrijpelijk.

Bij de jongens is het aantal zeer significante verschillen ditmaal niet opvallend groot. Bij de zoons van werkers in de agrarische sector, treffen we een aanzienlijk aantal verschillen aan, die niet de 1%, maar wel de 5%-drempel van betrouwbaarheid bereiken. Onder hen vinden we een voorkeur voor de technische handarbeid ($\phi = + .11$) en een afwending van de M-, TN- (beide $\phi = - .11$) en LG-schaal ($\phi = - .09$). De zeer bescheiden voorkeur voor agrarische arbeid, bereikt zelfs nog niet de 5%-drempel van significantie. Deze verschillen worden hier, ondanks hun dubieuze betrouwbaarheid, gereleveerd, omdat zij onderling consistent zijn. Er ligt een indicatie in besloten, dat de betreffende groep jongens sterker praktisch ingesteld is, dan de overigen.

Bij de jongens, waarvan de vader in de sectoren handel of verkeer werkzaam is, constateerden we een afwending van de agrarische arbeid ($\phi = - .13$). Tenslotte vonden we bij de subgroep vrije beroepen en onderwijs, gecombineerd met administratie een meer dan gemiddelde voorkeur voor de LG- en M-schalen (ϕ resp. $+ .12$ en $+ .15$). De betekenis van het eerstgenoemde verschil is niet duidelijk, die van de beide laatstgenoemde past weer bij de grote frequentie, waarmee in deze gecombineerde subgroepen het v.h.m.o. als voortgezet onderwijs zal worden gekozen.

Een analyse van de gecombineerde scores van jongens en meisjes kwam in dit geval niet in aanmerking, omdat de sexeverhouding in de subgroepen niet voldoende constant bleek te zijn.

Voor de samenhangen, die in deze paragraaf aan de orde kwamen, geldt dus in hoofdzaak hetzelfde als bij de verschillen, die verband houden met het beroepsniveau van de vader: zij zijn betrekkelijk klein in aantal en van geringe praktische betekenis, maar zover zij voorkomen, als regel wel begrijpelijk. We stuiten thans ook op samenhangen, die niet steeds zonder meer in verband kunnen worden gebracht met milieuverschillen in begaafdheidsniveau.

12.6. Onderwijsrichting

De laatste vraag, die ons zal bezighouden betreft de eventuele verschil-

len tussen de leerlingen van schooltypen, die tot verschillende „zuilen” behoren. Deze vraag is ook voor de ijking van belang, omdat we hebben geconstateerd, dat het protestants-christelijk onderwijs in te kleine getale in onze steekproef vertegenwoordigd was.

Vanzelfsprekend behoeft de richting van het onderwijs, die door een kind wordt bezocht, niet altijd te worden opgevat als een ondubbelzinnige indicatie aangaande de geloofsgroep, waartoe het gezin behoort. Wel kan men, op grond van dagelijkse observatie, het vermoeden uitspreken, dat slechts een zeer kleine minderheid van de leerlingen van het R.K.-onderwijs niet uit katholieke gezinnen komt. Maar voor het Prot. Chr. onderwijs geldt een analoge overweging zeker niet; een niet gering deel van de leerlingen komt uit niet-orthodox protestantse gezinnen. Het openbaar onderwijs op zijn beurt zal zijn leerlingen wel in meerderheid uit buitenkerkelijke of vrijzinnig-protestantse gezinnen recruterende, maar trekt toch ook een niet te verwaarlozen aantal kinderen uit de andere groeperingen. Deze overwegingen maken het bij voorbaat onwaarschijnlijk, dat we bij ons onderzoek aanzienlijke verschillen tussen de voorkeurspatronen van de leerlingen van het openbaar en protestants-christelijk onderwijs zullen vinden. Indien er al verschillen zijn, moeten we deze in de eerste plaats verwachten tussen het R.K. en het overig onderwijs. Bij oudere leeftijdsgroepen zijn daar inderdaad wel voorkeursverschillen aangetoond. In een Duits onderzoek uit de eerste jaren na 1900 (geciteerd bij ⁸⁰) werd gevonden, dat de beroepskeuze van de R.K. abituriënten een aanzienlijk andere frequentieverdeling vertoont, dan die van de protestanten. In hoofdzaak kan dit verschil worden getypeerd met de uitspraak, dat bij de R.K. de voorkeur voor de alfarichtingen verhoudingsgewijs groter was. Geheel analoge verschillen treden in ons hoger onderwijs aan het licht, zoals nog uit recente gegevens (³², ³⁴) blijkt.

Bij analyse van ons materiaal werden in eerste instantie inderdaad een aantal opmerkelijke verschillen gevonden, die op soortgelijke richtingsverschillen schenen te wijzen. Bij de R.K.-jongens bleek de voorkeur voor de TN-richting aanmerkelijk geringer dan bij de overige jongens ($\phi = -.20$). Bij de meisjes waren er veel meer verschillen: de R.K.-meisjes kozen significant meer de items uit de AD- ($\phi = +.34$), HA- ($\phi = +.31$), TH- ($\phi = +.23$) en TN-schaal ($\phi = +.20$) en significant minder de items uit de SO-schaal ($\phi = -.24$). Tussen de leerlingen van het openbaar en het protestants-christelijk onderwijs werden geen significante verschillen gevonden.

Deze bevindingen waren aanleiding, om de frequentieverdeling van de drie hoofdrichtingen, naar sexe gesplitst, meer in detail te onderzoeken. Daarbij bleek, dat bij vijf van de zes groepen de afwijkingen van het gemiddelde beeld slechts weinig te betekenen hadden. Bij de jongens bleken de leerlingen van het openbaar onderwijs een iets meer dan gemiddelde score op de TN- en LG- en dientengevolge ook op de M-schaal te behalen.

Bij de R.K.-jongens waren de TN- en M-score iets aan de lage kant, terwijl daarentegen de AV-, SO- en L-score iets boven het gemiddelde lagen. De numeriek kleine Prot. Chr.-groep vertoonde geen significante afwijkingen van het gemiddelde. Bij de vrouwelijke leerlingen van het openbaar onderwijs waren, behalve de SO-score, ook de gemiddelden op de AA-schaal en — in geringe mate — op de AV- en VB-schaal aan de hoge kant, terwijl de TH-, HA- en AD-scores iets te laag waren. Bij de Prot. Chr.-meisjes vielen vooral de iets aan de hoge kant liggende gemiddelden op de VB- en SO-schalen op. In al deze gevallen waren de verschillen echter numeriek van weinig betekenis: statistisch juist significant, of zelfs alleen maar als tendenties in het materiaal aanwezig. De enige groep, die een duidelijk afwijkend beeld bleef vertonen, was die van de R.K.-meisjes. In tabel 12.1 wordt van de frequentieverdeling van deze groep een overzicht gegeven.

TABEL 12.1. Belangstellingskarakteristiek van vrouwelijke leerlingen van het R.K.-onderwijs (N = 189)

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M
< D ₁	—	5	—	13	15	6	2	12	17	2	18
D ₂	—	20	19	18	27	10	5	35	46	16	29
D ₃	24	42	39	48	52	19	14	47	77	32	44
D ₄	—	57	—	61	70	33	32	67	101	51	61
D ₅	48	77	63	94	100	48	42	83	119	65	83
D ₆	73	107	76	105	114	64	67	106	139	96	103
D ₇	94	130	104	130	145	97	93	130	156	111	118
D ₈	107	146	132	153	155	125	126	150	173	130	148
D ₉	141	170	159	170	173	165	155	171	182	159	168
≥ D ₉	48	19	30	19	16	24	34	18	7	30	21

Er zijn duidelijk significante verschillen op de TH-, AD-, TN-, HA- en SO-schalen. Het algemene beeld van deze frequentieverdeling wordt dus gekarakteriseerd door een verschuiving in de richting van het masculiene beeld. Deze verschuiving treedt echter ook op bij de meisjes die v.g.l.o. gaan volgen of na de zesde klas de praktijk in gaan. Onderzocht zou dus moeten worden, of de gevonden correlatie wellicht met een verschil in bestemming samenhangt. Indien b.v. meer R.K.-meisjes de beide genoemde richtingen zouden kiezen, dan op grond van hun aantal in het totale materiaal te verwachten is, zou dit betekenen, dat we daarmee in onze conclusies rekening moeten houden. In dat geval doen zich in principe drie mogelijkheden voor: óf de karakteristieke trekken van tabel 12.1 moeten worden verklaard uit de grotere frequentie van v.g.l.o. en „praktijk”-meisjes, óf de karakteristiek van de beide bestemmingsgroepen is een gevolg van de aanwezigheid van een te grote concentratie van R.K.-meisjes in het betreffende gedeelte van ons materiaal, óf de beide kenmerken

versterken elkaar. Om dit te toetsen is eerst onderzocht, welke verschillen er eventueel bestaan tussen de bestemming van de R.K.-meisjes en die van het totale materiaal. Tabel 12.2 geeft een overzicht van de percentages, die naar de verschillende bestemmingen gaan.

TABEL 12.2. Frequentieverdeling der schoolbestemming bij meisjes (percentages)

	v.h.m.o.	m.u.l.o.	n.o.	v.g.l.o.	practijk	N
R.K.	5.8	30.7	18.0	35.4	7.4	189
Totaal	7.6	33.7	31.4	16.4	4.3	487

De percentages v.h.m.o.- en m.u.l.o.-aspiranten blijken onder de R.K.-meisjes dus iets kleiner te zijn dan in het totale materiaal, maar dit verschil heeft, omdat het numeriek slechts om kleine groepen gaat, weinig te betekenen. Daarentegen is het percentage, dat naar het n.o. gaat duidelijk veel te laag en het percentage, dat naar v.g.l.o. of de praktijk gaat, omgekeerd veel te hoog.

We hebben hier dus inderdaad te maken met een gecompliceerde samenhang. Bij de interpretatie van de bijzondere belangstellingskarakteristiek, die bij de R.K.-meisjes werd gevonden, moet rekening worden gehouden met het feit, dat onder hen een relatief zeer groot aantal rechtstreeks in de praktijk, of naar het v.g.l.o. gaat. Deze relatief sterke preferentie van R.K. meisjes voor het v.g.l.o. is niet slechts een toevallige eigenaardigheid van ons materiaal. Bij analyse van de landelijke cijfers over het schoolbezoek der 12- t.m. 15-jarigen komt dezelfde tendentie naar voren ⁽⁹⁾.

Om na te gaan, of de gevonden karakteristiek inderdaad met de richting samenhangt, doet men het beste, de meisjes, die naar v.g.l.o. of praktijk gaan, van het onderzoek uit te sluiten. Wij hebben daarom uit het totale materiaal de gegevens gelicht, die betrekking hadden op de meisjes, die naar v.h.m.o., m.u.l.o., huishoudschool of overbruggingsjaar gaan. In totaal betrof dit 374 leerlingen. Voor de schalen, waarop de R.K.-meisjes zich duidelijk van de anderen onderscheiden, is nu voor deze nieuwe steekproef onderzocht, of nog steeds richtingsverschillen voorkomen. Van de gebieden TH en TN is in tabel 12.3 de frequentieverdeling voor deze beide schalen, gesplitst naar de drie hoofdrichtingen weergegeven.

Uit deze tabel blijkt, dat de richtingsverschillen nu zijn weggefallen. Het iets hoger liggende gemiddelde van de R.K.-meisjes op de TN-schaal verschilt niet significant van dat van de anderen, ondanks het feit, dat de aantallen meisjes in dit geval toch reeds voldoende groot zijn, om betrekkelijk kleine verschillen aan het licht te brengen. Op de overige schalen blijken de verschillen in de ligging van de mediaan eveneens niet meer dan hoogstens één deciel te bedragen.

Op grond van deze bevinding mag o.i. worden geconcludeerd, dat er van een specifiek richtingsgebonden verschil in de belangstellingskarak-

TABEL 12.3. Scoreverdeling op de TB- en TN-schalen bij meisjes van verschillende richting.

	RK	TH PC	OL	RK	TN PC	OL
< D ₁	—	—	—	—	—	—
D ₂	—	—	—	9	1	12
D ₃	21	9	45	—	—	—
D ₄	—	—	—	20	10	46
D ₅	40	17	96	32	21	83
D ₆	56	26	133	42	28	115
D ₇	69	29	165	61	35	163
D ₈	77	34	189	77	38	186
D ₉	93	40	214	92	40	203
≥ D ₉	10	2	15	11	2	26

teristiek ook bij de meisjes niet kan worden gesproken. Indien er al verschillen zijn, zijn deze van zo bescheiden aard, dat zij voor de praktijk geen betekenis hebben. Eventuele verschillen in de voorkeur, die zich b.v. aan het einde van het middelbaar onderwijs zouden openbaren, moeten dus hun oorsprong hebben in factoren, die pas tijdens de puberteit hun werking ontplooien.

BIJLAGEN

B I J L A G E I
STATISTISCHE SIGNIFICANTIE VAN VOORKEURSVERSCHILLEN

	TH	AV	TN	VB	AA	HA	AD	LG	SO	L	M	n _g
1. <i>Bestemming</i>												
V.H.M.O.	--		++		--		++	++		--	++	65
M.U.L.O.	--						++			--	++	153
L.T.S.	++						--			++	--	185
V.G.L.O.				+								50
Practijk			--							+	--	29
Landbouw			--								--	29
Overg.jaar		--	++		++			--			++	25
V.H.M.O.			++		++	--	--	++		--	++	37
M.U.L.O.	--		--				--	++		--	++	164
L.N.O.			--				--	--	++	++	--	153
V.G.L.O.	++		+	++		++	++	--	--	++	--	80
Practijk	++					++	++			++	--	21
Overg.jaar	++				++	--	--	--		++	--	20
2. <i>Richting</i>												
L.T.S.												
Bankwerker	++	--	++		--		--			++		103
Instrumentmaker	++	--	++	--	--	--	--		--	++	++	51
Electromonteur	++	--	++	--	--		--		--	++	++	75
Loodgieter	++						--			--	--	32
Schilder	--	+	--			--	--	++	+	--	--	38
Straatmaker							--					21
Timmerman		+		--			--					32
Meubelmaker		++	--	+			--		+	++		27
Bakker	--	++	--	+			--			++		87
Zetter	--	++					--					21
Drukker		++		--			--	++				23
V.H.M.O. bevorderd na streng selectie												
jongens	--	--	++	--	--	--	--	++		--	++	83
meisjes			++	--	++	--	--	++		--	++	71
3. <i>Predictie V.H.M.O.</i>												
bevorderd na strenge selectie versus niet-bevorderd												
jongens			+	--	--	--	--			--	++	83/49
meisjes						--	--				++	71/33

Toelichting: Bij de significantieberekening is als regel gebruik gemaakt van t-toetsen. De significantiebepaling geschiedde in de rubrieken 1 en 2 ten opzichte van de schaalgemiddelden van het ijkingsmateriaal. Indien het gemiddelde van de speciale groep voor een bepaalde schaal significant lager was dan dat van het ijkingsmateriaal, is dit aangeduid met — ($P < .05$) of — — ($P < .01$). Een significant hoger gemiddelde van de speciale groep is aangeduid met + ($P < .05$) of ++ ($P < .01$).

In rubriek 1 is bij de berekening van de standaarddeviatie van het verschil rekening gehouden met het feit, dat de betreffende speciale groepen tevens deel uitmaken van het ijkingsmateriaal. In verband daarmee is een formule van MC NEMAR (97, pag. 100) gebruikt:

$$s_D = \frac{s}{\sqrt{n_s}} \sqrt{1 - \frac{n_s}{N}}$$

Hierin is s de standaarddeviatie op de betreffende schaal van het ijkingsmateriaal; N de omvang van het ijkingsmateriaal (531 jongens resp. 487 meisjes); n_s het in de laatste kolom van deze bijlage vermelde aantal personen in de speciale groep.

De significantiebepaling in rubriek 3 heeft betrekking op een vergelijking tussen degenen, die na strenge selectie op een v.h.m.o.-school werden bevorderd en degenen, die al dan niet na strenge selectie, aan het einde van het eerste jaar niet werden bevorderd. Een + teken wijst hier op een hogere gemiddelde score bij eerstgenoemde groep.

In een enkel geval van sterk van het normale afwijkende frequentieverdelingen, is een distributievrije toets (WILCOXON) gebruikt voor de bepaling der significantie.

B I J L A G E II

DETERMINEERTABEL VOOR DE RICHTINGKEUZE IN HET VOORTGEZET ONDERWIJS

Schaal	Variatie	Jongens	Meisjes
TH	+	instrumentmaker, electromonteur, bankwerker, loodgieter	v.g.l.o., praktijk
	—	v.h.m.o., m.u.l.o., schilder, bakker, typograaf	
AV	+	schilder, bakker, meubelmaker, typograaf, v.g.l.o., praktijk	
	—	instrumentmaker, electromonteur, v.h.m.o.	v.h.m.o., v.g.l.o.
TN	+	v.h.m.o., instrumentmaker, electromonteur, bankwerker	v.h.m.o., v.g.l.o.
	—	schilder, bakker, landbouwer	
VB	+	(banket)bakker, landbouwer v.g.l.o.	n.o.
	—	timmerman, meubelmaker, typograaf	v.h.m.o.
AA	+	landbouwer	v.h.m.o.
	—	instrumentmaker, electromonteur, bankwerker, v.h.m.o.	
HA	+	bankwerker	v.g.l.o., praktijk
	—		v.h.m.o.
AD	+		v.g.l.o.
	—	schilder, bakker, typograaf	v.h.m.o.
LG	+	v.h.m.o., schilder, typograaf, instrumentmaker	v.h.m.o., m.u.l.o.
SO	+	bakker, straatmaker, v.g.l.o. praktijk	n.o.
	—	v.h.m.o.	v.h.m.o.
L	+	bakker, bankwerker, typograaf, praktijk	n.o., praktijk
	—	v.h.m.o.	v.h.m.o.
M	+	v.h.m.o., instrumentmaker, electromonteur	v.h.m.o.
	—	loodgieter	

Toelichting: In de determineertabel zijn per schaal de richtingen vermeld, waarvan bij onderzoek werd gevonden, dat de mediale score significant van die van het ijkingsmateriaal verschilde. Opgenomen zijn alleen de gevallen, waarin de mediale score tenminste hoger dan het 60e percentiel (+ scores) of lager dan het 40e percentiel (— scores) was. Bij groepen, waarvan nog slechts een klein aantal personen is onderzocht, zijn de minimale afwijkingen van de doorsnee, in verband met de significantie-eis, echter aanzienlijk groter, dan de genoemde percentielwaarden.

ERRATUM

De op pagina 37 vermelde formule voor de toevalsrelatie moet luiden:

$$r = -m/108$$

De verwachte waarden voor de toevalsrelaties tussen de schalen variëren derhalve tussen $-.083$ en $-.167$.

BIJLAGE III
GRAFISCHE DETERMINEERTABEL

voortgezet onderwijs voor jongens

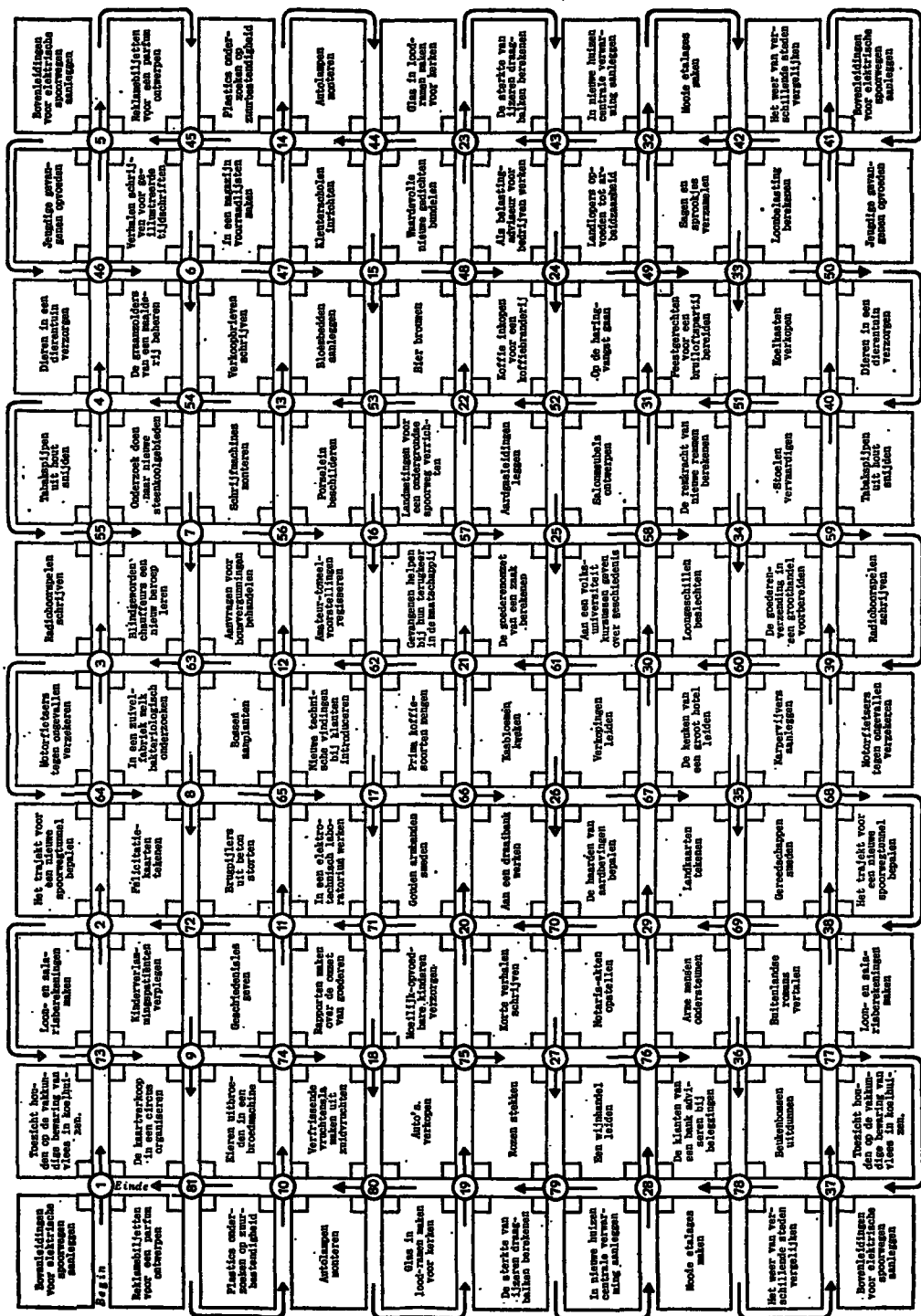
D₁ D₂ D₃ D₄ D₅ D₆ D₇ D₈ D₉

										TH										schilder
																				bakker
																				v.h.m.o.
																				m.u.l.o., typograaf
																				bankwerker, electromonteur, loodgieter
																				instrumentmaker
																				AV
																				v.h.m.o.
																				electromonteur, instrumentmaker
																				landbouw, bakker, praktijk, v.g.l.o.
																				timmerman
																				typograaf
																				meubelmaker
																				schilder
																				TN
																				landbouw, praktijk, schilder
																				bakker
																				bankwerker
																				electromonteur
																				instrumentmaker
																				v.h.m.o.
																				VB
																				timmerman
																				landbouw, praktijk, straatmaker, v.g.l.o.
																				bakker
																				AA
																				instrumentmaker
																				v.h.m.o.
																				bankwerker
																				landbouw
																				HA
																				schilder
																				AD
																				schilder
																				bakker, typograaf
																				LG
																				timmerman
																				schilder
																				v.h.m.o.
																				typograaf
																				SO
																				instrumentmaker
																				bakker
																				praktijk, straatmaker
																				L
																				v.h.m.o.
																				loodgieter, praktijk, typograaf
																				bankwerker, meubelmaker
																				bakker
																				M
																				loodgieter
																				electromonteur, instrumentmaker
																				v.h.m.o.

BIJLAGE V

Afbeelding van één pagina van de Beroepeninteressesettest

Het auteursrecht van deze test berust bij de uitgeversmaatschappij C. J. Hogrefe te Göttingen. Een Nederlandse editie zal vermoedelijk in de loop van 1959 verschijnen.



LITERATUURLIJST

1. AMMONS, R. B. A projective test for vocational research at the college level.
J. appl. Psychol. 34, 198-205 (1950).
2. AMMONS, R. B. Vocational Apperception Test.
Louisville 1949.
3. ANON Adreslijst van het lager onderwijs in Nederland.
Zutphen 1955.
4. ANON Comptes-rendus du 13e congrès de l'Association Internationale de Psychologie Appliquée, Rome 1958.
In voorbereiding.
5. ANON De grafische beroepen, een psychologische analyse.
Amsterdam 1949.
6. ANON Groninger Afsluitingsonderzoek Lager Onderwijs G A L O. Uitgave van het Groninger Instituut voor Toegepaste Psychologie.
Groningen z.j.
7. ANON Motivation Indicator.
J. consult. Psychol. 11, 157 (1947).
8. ANON Rapport over een onderzoek naar de stand van het Gewoon Lager Onderwijs in Noord-Brabant.
Den Bosch 1957.
9. ANON De schoolsituatie der 12 t/m 15-jarigen in Nederland, kwantitatief beschouwd.
Uitgave Kath. Sociaal-Kerkelijk Instituut, Den Haag 1958.
10. ANON Selection for grammar schools.
Publicatie van de National Foundation for Educational Research in England and Wales.
In voorbereiding.
11. ANON Sequential Test of Educational Progress S T E P. Uitg. Educational Testing Service.
Princeton 1957.
12. ANON De stand van de onderwijs-research in Nederland.
Publicatie nr. 4 van de Katholieke Centrale voor Studie en Research.
Tilburg 1959.
13. ANON Technical recommendations for psychological tests and diagnostic techniques.
Supplement to the Psychological Bulletin, 51, nr. 2, deel 2 (1954).

14. BALDWIN, G. B. Motivation Indicator.
Minneapolis 1946.
15. BARENDREGT, J. T. Een experimenteel-statistisch onderzoek naar de
waarde van klinische predictie.
Ned. Tijdschr. Psychol. 13, 425-442 (1958).
16. BAUMGARTEN, F. Die Charakterprüfung der Berufsanwärter.
Zürich 1946.
17. BAUMGARTEN, F. Evokative Tests. In: Stern, E. (ed.)
Die Tests in der klinischen Psychologie I,
396-418.
Zürich 1954.
18. BEMELMANS, F. Test projectif d'intérêts professionnels.
Schw. Z. Psychol. 12, 283-294 (1953).
19. BENNETT, G. K. e.a. Differential Aptitude Tests.
Uitg. Psychological Corporation.
New York 1947.
20. BERDIE, R. F. Factors related to vocational interests.
Psychological Bulletin 43, 137-157 (1944).
21. BOON VAN OSTADE, A. H. Itemanalyse ten behoeve van de validiteit met
behulp van de homogeniteitsindex van LONG en
LOEVINGER.
Niet gepubliceerde scriptie R.U. Leiden, 1958.
22. BORDEWIJK, W. Een onderzoek naar de factoren, die beroeps-
wens en beroepsvoorstelling van de proletari-
sche jeugd beïnvloeden.
Groningen 1949.
23. BRAINARD, P. P. en Occupational Preference Inventory.
BRAINARD, P. T. New York 1945.
24. BROUWER, W. H. Selectie en schoolsucces.
Groningen 1951.
25. BUROS, O. K. The 1940 Mental Measurements Yearbook.
Arlington 1941.
26. BUROS, O. K. The third Mental Measurements Yearbook.
Highland Park 1949.
27. BUROS, O. K. The fourth Mental Measurements Yearbook.
Highland Park 1953.
28. CARTER, H. D. The development of interests.
Yearbook Nat. Soc. Study of Educ. 43, 255-276
(1944).
29. CARTER, H. D. Resources for the consultant: the development
of vocational attitudes.
J. consult. Psychol. 4, 185-191 (1940)
30. C.B.S. Bevolking der gemeenten van Nederland op
1 Januari 1958.
Utrecht 1958.
31. C.B.S. De Nederlandse jeugd en haar onderwijs.
1953/54.
Utrecht 1955.

32. C.B.S. Statistiek der academisch gevormden.
12e Volkstelling, deel A 5.
Den Haag 1950.
33. C.B.S. Statistiek van het gewoon lager en voortgezet
gewoon lager onderwijs 1956.
Utrecht 1957.
34. C.B.S. Statistiek van het hoger onderwijs 1953/54.
Utrecht 1954.
35. C.B.S. Statistiek van het v.h.m.o. 1954/1955.
Utrecht 1956.
36. C.B.S. Statistiek van de voorlichting bij beroepskeuze
en personeelselectie 1957.
Utrecht 1958.
37. CHAUDAGNE, H. Un questionnaire d'intérêts pour garçons de 13
à 16 ans.
B I N O P 1952, 218-223.
38. CHORUS, A. Het intelligentie-onderzoek en zijn kwalitatieve
verdieping.
Utrecht 1947.
39. CHRISTIAENS, X. De betekenis en waarde van de Kuder belang-
stellingstest.
Tijdschr. Studie- en Beroepsoriënt. 4, 13-28
(1957).
40. CRONBACH, L. J. en
MEEHL, P. E. Construct validity in psychological tests.
In: Feigl, H. en Scriven, M. (ed.) The founda-
tions of science and the concepts of psychology
and psychoanalysis, 174-204.
Minneapolis 1956.
- 40a. CURETON, E. E. The definition and estimation of test reliability.
Educ. Psychol. Measurement 18, 715-738
(1957).
41. DARLEY, J. G. en
HAGENAH, TH. Vocational interest measurement.
Minneapolis 1955.
42. DAVIS, F. B. The standard error of measurement of a dif-
ference between two obtained scores.
Stencil G.U., Amsterdam 1957/58.
43. DERIVIÈRE, R. Le contrôle des intérêts professionnels dans
l'enseignement secondaire.
Travail humain 16, 1-30 (1953).
44. DEWEY, J. Interest and effort in education.
Boston 1913.
45. DUFF, J. F. en
THOMSON, G. H. Social and geographical distribution of intel-
ligence in Northumberland.
Brit. J. Psychol. 14, 192-198 (1923).
46. EEKELEN, W. F. VAN Beroepsvoorkeur als hulpmiddel voor de psy-
chologische diagnostiek.
Niet gepubliceerde scriptie, Gemeentelijke
Universiteit Amsterdam, z.j.
47. FESTINGER, L. en
KATZ, D. (ed.) Research methods in the behavioral sciences.
New York 1953.

48. FITZPATRICK, T. P. en WISEMAN, S. An interest test for use in selection for technical education.
Brit. J. Educ. Psychol. 24, 99-105 (1954).
49. FRYER, D. The measurement of interests in relation to human adjustment.
New York 1931.
50. GIESSEN, R. W. VAN DER Enkele aspecten van het probleem der predictie in de psychologie.
Amsterdam 1957.
51. GINZBERG, E. e.a. Occupational choice.
New York 1951.
52. GOODENOUGH, F. Mental testing.
New York 1949.
53. GREENE, E. B. Michigan Vocabulary Profile Test.
Yonkers-on-Hudson 1939.
54. GROOT, A. D. DE De betekenis van „significantie” bij verschillende typen onderzoek.
Ned. Tijdschr. Psychol. 11, 398-409 (1956).
55. GROOT, A. D. DE Psychologie als wetenschap. I, Het empirisch-wetenschappelijk denken.
In voorbereiding.
56. GROOT, A. D. DE Genormaliseerde werkclassificatie.
Mens en Onderneming 7, 401-415 (1953)
57. GUILFORD, J. P. Fundamental statistics in psychology and education.
New York 1950.
- 57a. GUILFORD, J. P. A simple scoring weight for test items and its reliability.
Psychometrika 6, 367-374 (1941).
58. GULLIKSEN, H. Theory of mental tests.
New York 1949.
59. HAMILTON, M. A simple diagram for obtaining tetrachoric correlation-coefficients.
Brit. J. Psychol., general section, 39, 168-171 (1948/49).
60. HARRIMAN, Ph. L. (ed.) Encyclopedia of psychology.
New York 1946.
61. HEYDEN, Ph. M. VAN DER Begaafdheid en beroep.
Groningen 1953.
62. HOEL, P. G. Introduction to mathematical statistics.
New York 1947.
63. HOFSTÄTTER, P. Psychologie.
Frankfurt am Main, 1957.
64. HORNSTRA, L. De mens in de organisatie van de arbeid.
Amsterdam 1951.
65. HOYT, C. Test reliability obtained by analysis of variance.
Psychometrika 6, 153-160 (1941).
66. IRLE, M. Berufs Interessen Test.
Göttingen 1955.

67. IRLE, M. Handanweisung zum Berufs Interessen Test. Göttingen 1955.
68. IRLE, M. Eine Methode zur Bestimmung beruflicher Interessen und erste Ergebnisse ihrer Anwendung. Diss. Göttingen 1955.
69. JAMES, W. The principles of psychology. London 1891.
70. JASPEN, N. Serial correlation. Psychometrika 11, 23-30 (1946).
71. JONGE, H. DE Medische statistiek I. Leiden 1958.
72. KELLEY, T. L. Educational guidance. Columbia Teachers College Contributions to Education. New York 1914.
73. KELLEY, T. L. Interpretation of educational measurement. Yonkers-on-Hudson 1927.
74. KELLEY, T. L. The scoring of alternative responses with reference to some criterion. J. Educ. Psychol. 25, 504-510 (1934).
75. KELLY, E. L. en FISKE, D. W. The prediction of success in the VA training program in clinical psychology. Amer. Psychologist 5, 395-406 (1950).
76. KENDALL, M. G. Rank correlation methods. London 1948.
77. KOUWER, B. J. Hoe kiezen wij onze testmethoden? Ned. Tijdschr. Psychol. 12, 243-251 (1957).
78. KRETCH, D. en CRUTCHFIELD, R. S. Theory and problems of social psychology. New York 1948.
79. KUDER, F. D. A comparative study of some methods of developing occupational keys. Educ. Psychol. Measurement 17, 105-114 (1957).
80. KUDER, F. D. Preference Record; occupational. Chicago 1956.
81. KUDER, F. D. Preference Record; occupational, research handbook. Chicago 1956.
82. KUDER, F. D. Preference Record; personal. Chicago 1954.
83. KUDER, F. D. Preference Record; vocational. Chicago 1939.
- 83a. KUDER, F. D. Manual to the Preference Record; vocational. Chicago 1946.
84. LANGEVELD, M. J. Inleiding tot de studie der paedagogische psychologie. Groningen 1937.
85. LARCEBEAU, S. Un questionnaire d'intérêts pour garçons. BINOP 1955, 74-87.

86. LAZARSFELD, P. Jugend und Beruf.
Jena 1931.
87. LEERTOUWER, C. Onderzoek naar de stand van het Lager On-
derwijs in Noord-Brabant.
Paed. Studiën 35, 557-564 (1958).
88. LEHMAN, H. C. en One more study of permanence of interests.
WITTY, P. A. J. educ. Psychol. 22, 481-492 (1931).
89. LEVINE, F. R. en Adolescent vocational interests and later occup-
WALLEN, R. ation.
J. appl. Psych. 38, 428-431 (1954).
90. LEWIN, K. Field theory in social science.
New York 1951.
91. LINDQUIST, E. F. (ed.) Educational measurement.
Washington 1951.
92. LOEVINGER, J. A systematic approach to the construction and
evaluation of tests of ability.
Washington 1947 (serie Psychological Mono-
graphs).
93. LONG, J. A. Improved overlapping methods for determining
the validities of test items.
J. exper. Educ. 2, 264-268 (1934).
94. LORD, F. M. Estimating test reliability.
Educ. Psychol. Measurement 15, 325-336 (1955)
95. MCCLELLAND, D. C. e.a. The achievement motive.
New York 1953.
96. MCKENZIE, R. M. An occupational classification for use in vocat-
ional guidance.
Occupational psychology 28, 108-117 (1954).
97. MCNEMAR, Q. Psychological statistics.
New York 1955.
98. MEEHL, P. E. Clinical versus statistical prediction.
Minneapolis 1954.
99. MEEHL, P. E. Wanted — a good cookbook.
Amer. Psychologist, 11, 263-272 (1956).
100. MONROE, W. S. (ed.) Encyclopedia of educational research.
New York 1950.
101. MOSER, U. Psychologie der Arbeitswahl und der Arbeits-
störungen.
Bern 1953.
102. NASS, C. A. G. Een Monte-Carlo-bepaling van overschrijdings-
kansen in verband met een keuzetest.
In voorbereiding.
103. NAVILLE, P. La cristallisation de l'illusion professionnelle.
J. Psychol. norm. et pathol. 46, 316-348, (1953).
- 103a. NOLL, V. H. An introduction to educational measurement.
Boston (1957).
104. PARSONS, F. Choosing a vocation.
London 1909.

105. PEAK, H. Problems of objective observation. *In: Festinger, L. en Katz, D. Research methods in the behavioral sciences*, 243-299. New York 1953.
106. PERRY, D. K. Forced-choice versus L I D response items in vocational interest measurement. *J. appl. Psychol.* 39, 256-262 (1955).
107. RAPAPORT, D. e.a. Diagnostic Psychological Testing. Chicago 1945.
108. RENNES, P. Construction et étude d'un nouveau questionnaire d'intérêts. *Travail Humain* 15, 41-55 (1952).
109. ROCK, R. T. en WESMAN, A. The comparative efficiency of various methods of weighting test items. *Psychological Bulletin* 36, 569-570 (1939).
110. ROSENBERG, N. Stability and maturation of Kuder interest patterns. *Educ. Psychol. Measurement* 13, 449-458 (1953).
111. Rijksarbeidsbureau Classificatie van de beroepen naar hun onderlinge verwantschap. Den Haag 1952.
112. SIEGEL, S. Non-parametric statistics. New York 1956.
113. SPRANGER, E. Lebensformen. Halle an der Saale 1925.
114. STEFFLRE, B. The reading difficulty of interest inventories. *Occupations* 26, 95-96 (1947).
115. STELLWAG, H. W. F. Selectie en selectiemethoden. Groningen 1955.
116. STERN, W. Allgemeine Psychologie. Den Haag 1935.
117. STINISSEN, J. Collectieve intelligentietest. Antwerpen 1955.
118. STOUFFER, S. A. e.a. Measurement and prediction (deel IV van de serie: Studies in social psychology in World War II). Princeton 1950.
119. STOUFFER, S. A. e.a. Studies in social psychology in World War II. 4 delen. Princeton 1949 e.v.
120. STRELLER, J. Wörterbuch der Berufe. Stuttgart 1953.
121. STRONG, E. K. Manual for the Vocational Interest Blank for Men. Stanford 1951.
122. STRONG, E. K. Vocational Interest Blank for Men. Stanford 1938.

123. STRONG, E. K. Vocational Interest Blank for Women. Stanford 1946.
124. STRONG, E. K. Vocational interests of men and women. Stanford 1943.
125. STRONG, E. K. Vocational interest 18 years after college. Minneapolis 1955.
126. SUPER, D. E. Appraising vocational fitness. New York 1949.
127. SUPER, D. E. The psychology of careers. New York 1957.
128. TAYLOR, K. VON F. Reliability and permanence of vocational interests of adolescents. J. exp. Educ. 11, 81-87 (1942).
129. THORNDIKE, R. L. en LORGE, I. Lorge-Thorndike Intelligence Tests. Boston 1957.
130. TOBI, E. J. en LUYCKX, A. Herkomst en toekomst van de middenstander. Amsterdam 1950.
131. TRAVERSIN, A. M. Valeur et place de l'intérêt dans la psychologie appliquée. In voorbereiding.
- 131a. TRYON, R. C. Reliability and domain validity: reformulation and historical critique. Psychol. Bull. 54, 229-249 (1957).
132. TYLER, L. E. A comparison of the interests of english and american school-children. J. genet. Psych. 88, 175-181 (1956).
133. TYLER, L. E. The development of vocational interests I. J. genet. Psychol. 86, 33-44 (1955).
134. Unesco World Survey of Education. Paris 1955.
135. VUYK, R. De waarde van de SZONDI-test. Ned. Tijdschr. Psychol. 11, 447-468 (1956).
136. WATERINK, J. Over opmerkzaamheid en opmerkzaamheidsconcentratie. Ned. Tijdschr. Psychol. 1, 97-125 (1946).
137. WECHSLER, D. The measurement of adult intelligence. Baltimore 1944.
138. WIEGERSMA, S. Leesvaardigheidstest voor het onderzoek van de mechanische leesvaardigheid. Leiden 1958.
139. WIEGERSMA, S. Een onderzoek naar de geldigheid van de SZONDI-test voor de psychologische praktijk. Ned. Tijdschr. Psychol. 5, 108-125 (1950).
140. WIEGERSMA, S. De selectie bij het nijverheidsonderwijs voor jongens. Het Nijverheidsonderwijs 1957, 736-741 en 761-765.

141. WIEGERSMA, S. en
HALFFMANN, J.
De waarde van een schriftelijke testserie voor
selectie bij de toelating tot het lager technisch
onderwijs.
Paed. Stud. 32, 257-273 (1955).
142. WYMAN, J. B.
Tests of intellectual, social and activity interests
in Terman, L. M. Genetic Studies of genius I.
ch. XVI.
Stanford 1925.
143. ZUTHEM, H. J. VAN
Beroepsclassificatie en beroepsstratificatie.
Beroepskeuze 3, 34-41 (1956).

SAMENVATTING

Inleiding. Ten dienste van het belangstellingsonderzoek bij de differentiatie na de lagere school, is de *Berufs Interesses Test* (B I T) van Dr. M. IRLE, een test van het KUDER-type, voor Nederland bewerkt. Deze test, die oorspronkelijk bedoeld was voor het onderzoek van een enigszins oudere leeftijdsgroep, is vervolgens op zijn psychometrische kwaliteiten speciaal in verband met het onderzoek van ca. 12-jarige kinderen getoetst. Daarna is de B I T gebruikt voor een studie van de belangstellingskarakteristiek van verschillende groepen kinderen, die, in verband met de richtingkeuze na de lagere school, kunnen worden onderscheiden.

Hoofdstuk I. Belangstelling wordt hier gedefinieerd als: een aspect van de motivatie, dat bestaat uit de bewust-beleefbare, (semi-)permanente gerichtheid op bepaalde categorieën emotioneel stimulerende en tot actie uitlokkende situaties.

Gewezen wordt op de mogelijkheid, de ontwikkelingsgang der interesses te interpreteren in termen van de Lewiniaanse ontwikkelingstheorie. Ook de opvattingen van enkele andere auteurs over de genese der interesses worden besproken. Hoewel zonder meer duidelijk is, dat de interesses een ontwikkeling doormaken, blijkt het niet gemakkelijk een scherp omlijnde karakteristiek van de specifieke interesses van een bepaalde leeftijdsfase te geven. Aan de hand van SUPER's onderscheiding in „manifest, expressed en inventoried interests”, wordt er op gewezen, dat de laatstgenoemde interesses de beste perspectieven bieden, om tot een systematische belangstellingskarakteristiek te komen.

Een bijzonder probleem vormt in ons geval de relatief geringe stabiliteit der interesses bij de 12- à 14-jarigen. Er wordt op gewezen, dat het desondanks aanbeveling verdient, bij de voorlichting bij school- en beroepskeuze rekening te houden met de actueel aanwezige motiverende krachten.

Hoofdstuk II. Na een bespreking van enkele hoofdpunten van de geschiedenis van het belangstellingsonderzoek en met name ook van de recente ontwikkeling op dit gebied in Europa, volgt een behandeling van een drietal hoofdvormen van belangstellingsonderzoek: beroepenlijsten, STRONG Vocational Interest Blank en KUDER Preference Record. De beroepenlijsten zijn in ons land nog veelvuldig in gebruik; gewezen wordt op de grote bezwaren, die aan de huidige toepassing als test zijn verbonden. Tegenover de vele merites van de STRONG-test staan enkele praktische nadelen, die een toepassing op grote schaal in Nederland momenteel niet mogelijk maken. Daarentegen biedt de KUDER-test wel gunstige perspec-

tieven. Een test van het KUDER-type is daarom voor de Nederlandse situatie bewerkt: de beroepeninteressesettest van IRLE. Van het principe van deze test en van de arbeid, die door IRLE bij de constructie en beproeving is verricht, wordt in de laatste paragraaf van dit hoofdstuk een overzicht gegeven.

Hoofdstuk III. De Nederlandse en Duitse tekst van de B I T worden naast elkaar gegeven. Een aantal principes van de vertaling worden toegelicht. Vervolgens wordt een handleiding voor de testinstructie gegeven.

Hoofdstuk IV. In de Nederlandse situatie is de vraag naar de mogelijkheid, een differentiatie naar het niveau van het te kiezen voortgezet onderwijs te maken, van grote praktische betekenis. Er wordt op gewezen, dat de differentiatie naar niveau tevens een differentiatie naar richting i.v.m. de beroepskeuze impliceert. Alleen wanneer men zich hiervan bewust is, kan men op verantwoorde wijze het kwantitatieve niveaubegrip hanteren. In verband met de B I T wordt de vraag naar de mogelijkheid van de constructie van niveauschalen gesteld. In de tekst worden twee procedures beschreven, die zijn toegepast bij de constructie van dergelijke schalen. Resultaat van de arbeid is in de eerste plaats een tweetal schalen, aangeduid als L (activiteiten, die geprefereerd worden door degenen, die lager voortgezet onderwijs kiezen) en M (activiteiten, die geprefereerd worden door degenen, die middelbaar voortgezet onderwijs kiezen). Een eerste validatiestudie betreffende deze schalen wordt beschreven. Bovendien wordt de constructie van een gewogen MO-schaal behandeld. In samenhang hiermee wordt het probleem van de weging van testitems besproken.

De correlatie van de schaalscores met het intelligentieniveau blijkt als regel zeer gering te zijn.

Hoofdstuk V. De psychometrische kwaliteiten van de negen testschalen, die door IRLE werden ontwikkeld en van de beide door ons ontwikkelde schalen L en M, zijn met het oog op het gebruik van de test in de zesde klas van de lagere school uitvoerig onderzocht. De homogeniteit is getoetst met behulp van de procedure van LONG-LOEVINGER. In 93% van de gevallen bleek de homogeniteit der items met hun schaal goed te zijn. Ten aanzien van één schaal, betrekking hebbende op de interesse voor commerciële arbeid, was een revisie gewenst. Het bleek mogelijk, een schaal met meer homogene items te construeren. Gegevens worden verder gerapporteerd over de intercorrelaties der testschalen en over hun betrouwbaarheid volgens de „split-halves”- en de KUDER-RICHARDSON-methode.

Hoofdstuk VI. Ten behoeve van de ijking van de test is een aselechte steekproef uit de bevolking der zesde klassen van lagere scholen gebruikt. De wijze van samenstelling van de steekproef en de contrôle op representativiteit worden beschreven. De test is voor jongens en meisjes afzonder-

lijk geijkt. Behandeld worden voorts de vraag naar de significantie van verschillen tussen schaalscores en de mogelijkheid tot het herkennen van toevalsscores.

Hoofdstuk VII. Inleidend tot het gebruik van de test voor het onderzoek van de belangstellingskarakteristiek wordt een aantal aspecten van het validiteitsprobleem ter discussie gesteld. Een belangrijk punt vormt in dit verband de aangelegenheid van de operationele definitie van het interessebegrip. Er wordt op gewezen, dat via het z.g. „Münchhausen-effect” een brug met de in hoofdstuk I gegeven begripsbepaling kan worden geslagen. In aansluiting aan DE GROOT, wordt er op gewezen, dat de begripsvaliditeit niet wordt gemeten, maar beoordeeld op grond van de uitkomsten van onderzoeken waarin het begrip of de test betrokken is. Gezien de uitkomsten van de in de volgende hoofdstukken te rapporteren studies, kan hier worden gehandeld alsof de begripsvaliditeit in voldoende mate bestaat.

Hoofdstuk VIII. Zoals ook in andere belangstellingstests blijken de sexeverschillen in de B I T zeer geprononceerd. De verwantschap met de KUDER-test geeft een mogelijkheid tot congruente validatie van de sexe-karakteristiek. Verder worden in dit hoofdstuk de sexeverschillen in de belangstelling meer in detail besproken.

Hoofdstuk IX. Het vraagstuk van de relatie tussen interesseltestscore en schoolkeuze in het voortgezet onderwijs, is onderzocht aan de hand van de resultaten van ruim 1000 zesde klassers, waarvan de bestemming na de lagere school reeds bekend was. De aanstaande v.h.m.o.-leerlingen, jongens zowel als meisjes, leveren een zeer typerend gemiddeld belangstellingspatroon; dit in tegenstelling tot de aanstaande m.u.l.o.-leerlingen. Belangwekkend zijn de verschillen, die tussen de aanstaande leerlingen van het nijverheidsonderwijs en degenen, die v.g.l.o. kiezen of na de lagere school de praktijk in gaan, blijken te bestaan. Vooral bij meisjes zijn de verschillen zeer geprononceerd, de uitkomsten van het onderzoek wijzen op het bestaan van een niet onbelangrijke categorie meisjes, die met technisch voortgezet onderwijs gediend zouden zijn. Ook bij degenen, die een agrarisch beroep kiezen, wordt een specifiek belangstellingspatroon gevonden. Tenslotte wordt nog een vergelijking gemaakt tussen degenen, die na de zesde klasse naar het v.h.m.o. gaan en de kinderen, die eerst een „brugklasse” bezoeken.

Hoofdstuk X. De verschillen, die er tussen de aspiranten voor verschillende vakopleidingen bestaan, zijn bij leerlingen van het lager technisch onderwijs onderzocht. Er blijken in de eerste plaats een tweetal ontwikkelingsreeksen te bestaan, gekenmerkt door het optreden van een aantal onderling consistente belangstellingskarakteristieken, die bij rangschikking van de opleiding steeds meer geprononceerd naar voren komen.

Dit zijn loodgieter — bankwerker — electromonteur — instrumentmaker en straatmaker — timmerman — meubelmaker — schilder. Leerling-(banket)bakkers hebben een zeer specifieke belangstellingskarakteristiek. Tenslotte blijkt, dat er tussen leerling-zetters en -drukkers geen verschillen in belangstellingskarakteristiek bestaan.

Een vergelijking met Duitse gegevens, die op oudere groepen betrekking hebben, toont een zeer grote mate van overeenstemming aan. Het is echter ook mogelijk, enkele leeftijdsverschillen aan te tonen, die onafhankelijk van de richtingkeuze voorkomen.

Hoofdstuk XI. Aan de hand van de studie van ruim 500 kinderen, leerlingen van eerste klassen van v.h.m.o.-scholen, is een onderzoek ingesteld naar de predictieve validiteit van de B I T. Aangetoond kan worden, dat bepaalde schalen inderdaad een predictieve waarde hebben. Het effect wordt duidelijker merkbaar, naarmate in de eerste klas een scherpere beoordeling plaatsvindt.

Hoofdstuk XII. Afsluitend is nog een onderzoek gedaan naar de correlatie tussen B I T-scores en een aantal personalia: leeftijd in de zesde klasse, woongebied, beroepsniveau vader, beroepssector vader en richting („zuil”) van het onderwijs. In een klein aantal gevallen kunnen correlaties van zeer bescheiden aard worden geconstateerd, die dikwijls verband schijnen te houden met intelligentieverschillen. Een bijzonder geval deed zich voor bij het onderzoek naar de correlatie tussen onderwijsrichting en belangstellingskarakteristiek, waar in één geval wel een duidelijke correlatie scheen te bestaan. Deze bleek echter te zijn veroorzaakt door een interveniërende variabele.

SUMMARY

The subject of this book is interest testing, especially with regard to the selection for secondary education in the Netherlands. Elementary schools in this country have six grades, after which the pupil has to choose from several possibilities for secondary education, which are on three levels. On the upper level there are grammar schools, which give a five- or six-year preparation for university entrance. At the intermediate level there are schools that give a four-year preparation for entrance into vocational training colleges. At the lower level there are trade schools, with a two to four years' programme, giving a combination of general education and prevocational training. Also at the lower level there are schools that give a continuation of elementary education for two more years, intended especially for pupils, who mean to leave school as early as is permitted. For further details on this educational system the reader is referred to a UNESCO publication (184).

The choice of the right type of secondary education is of great practical importance, because the level that one attains in occupational life is clearly correlated with the level of education one receives. Therefore many parents try to make their child enter into as high a level of secondary education as possible. This, however, results in many failures, which in their turn not only cause momentary setbacks, but may mean a handicap throughout life, because the time to enter into the right type of school has passed. Psychologists in the Netherlands therefore try to develop a system of psychological selection, which can be used in educational guidance.

Up to now, however, scant attention has been given to the possibilities that interest testing may present in this connection. Not a single well-known interest-test has as yet been adapted for use in the Netherlands, nor has any good test of Dutch origin been published. The only instruments available are occupational checklists, of very dubious value. To remedy this situation an interest-test of the KUDER-type, the *Berufs Interessen Test*, developed in Germany by Martin IRLE (66), was translated into Dutch. This test has been developed for use with a somewhat older age-group than that of the elementary school-leavers, who are usually twelve or thirteen years, or sometimes fourteen years of age. It was therefore necessary to investigate the psychometric qualities of this test for use with these comparatively young children. Moreover, as very few data have been published on the vocational interests of this age-group, it was necessary to collect data on this subject. The test therefore was used in the second part

of this study, to examine several groups that are distinguished in secondary education, with a view to establish the characteristic patterns of their interests.

A summary of the contents of this study will now be given chapter by chapter, and usually also section by section.

Chapter I. General remarks on the concept of interest

After a consideration of various definitions given by other authors, interests are defined as a conscious aspect of motivation, consisting of the (semi-)permanent attitudes towards certain categories of situations that stimulate affective experiences and invite actions. The concept is explicitly differentiated both from instinctive drives and depth-psychological motives.

The discussion following this definition mainly stresses some developmental aspects. It is pointed out i.a. that little is known of the characteristic interests of children in early adolescence, which is thought to be primarily due to the lack of a common frame of reference in describing interests. The distinction in manifest, expressed and inventoried interests is discussed; it is assumed, that the latter offer the best opportunities for studying interest, owing to their greater reliability and validity and because they seem to afford the best opportunity for coming to an agreement on the frame of reference.

The third section deals with the problem of stability of interests, especially in relation to age. A review of some findings reported in literature is given. It is pointed out that on the data available it seems improbable that test results on interests, obtained in early adolescence, will have a long-term validity. However, this is no argument against using interest-test data in giving educational guidance to these children, as long as it can be proved that these data have concurrent and if possible also predictive validity.

Chapter II. The psychological examination of interests

The chapter begins with a short review of the history of interest testing, which highlights a few of the main events in this sector in America. The development of interest testing in Europe, which started only after World War II is discussed somewhat more in detail. It is pointed out that tests of the KUDER-type are usually appreciated more in Europe than those of the STRONG-type, while the reverse seems to be true for the United States. The second section reviews three instruments for interest testing, which are rather popular, either in the psychological world in general, or in the Netherlands. As an example of the latter the occupational checklist is chosen. The defects of this instrument, especially if used without an adequate statistical basis, are indicated. It is concluded that this instrument may be useful for eliciting interview data, or for forming a clinical judge-

ment, but that it is not a test in the strict sense of the word. However, it is possible to develop an adequate interest-test on the basis of an occupational checklist. The STRONG V I B is mentioned as one proof of this statement; also some, as yet unfinished, Dutch attempts to remedy the defects of this instrument are referred to.

The merits of the SRONG Vocational Interest Blank are pointed out. However, this test seems to be rather difficult for sixth-grade children. An experiment with a preliminary translation has shown that these children need about two hours on the average to complete the test. The scoring-system, too, is thought to be a handicap, because it makes expensive machine-scoring necessary. For these reasons no attempt has been made to adapt the STRONG V I B to the Dutch situation.

The KUDER Preference Record on the other hand has many attractive features, which seem to make it a good test for use in our situation. Adaptation of this test was seriously considered. Finally, however, it was decided to use the related vocational interest test of IRLE (B I T), which offers some slight advantages, especially in the way of scoring.

Section 3 discusses the work done by IRLE, in developing this test. Starting-point was a review by SUPER (¹²⁶, page 380), who pointed out that the same interests are used in quite a number of important tests. Starting with SUPER's eight common interests, IRLE, after eliminating the musical interest and adding agricultural (outdoor) and catering interests, constructed a test consisting of nine scales. Each one is designated by a name and an abbreviation, consisting of two letters that is used consistently in this book.

The scales are:

1. Mechanical trade (TH), comprising metal-working and construction;
2. Applied art (AV), related to the artistic interests in American tests, but somewhat more restricted in scope;
3. Science and engineering (TN);
4. Catering (VB), including all kinds of activities concerning the preparation and distribution of foods and drinks;
5. Agriculture and forestry (AA);
6. Persuasion (HA); including all kinds of commercial activities;
7. Clerical work (AD), including computational activities;
8. Literature and history (LG), the latter representing the „geisteswissenschaftliche“ aspect;
9. Social work and education (SO).

In constructing his test IRLE used an idea, developed by BALDWIN (¹⁴). The basic principle is best illustrated by a part of the test reproduced in appendix V. The test items are arranged in "streets"; at each "cross-road" one item must be chosen as most attractive. The choice is indicated by a cross in the tiny square in the corner of the preferred item-box, next to

the "cross-road". In total 162 choices must be made. The arrangement of the items is such that each of them is put to choice four times. There are 18 items in each scale, arranged in accordance with the pattern presented on page 34. The raw score on a scale can therefore vary between 0 and 72, with the provision that the sum of the scale-scores must be equal to 162. It can be shown that the last-mentioned restriction causes a slight negative correlation between the scales, which may, however, in the first analysis, be neglected, because of its small size. IRLE has given considerable attention to the homogeneity of the test items to their respective scales and also to the reliability of the scales. IRLE's data on homogeneity, reliability and intercorrelations of the scales are reported in this section. He has also made a study of 26 occupational categories; the relevant data are also reported.

Chapter III. The Vocational Interest Test

This chapter presents the German text of IRLE's test, side by side with the Dutch translation. Some remarks are made on the problems encountered in doing this translation. The test-instruction is described in detail.

Chapter IV. Educational level differences in vocational interest

As one of the main problems in selection for secondary education is the choice of the right level of education, it was thought worth while to try to augment the test scales by an educational level scale. In discussing this level-problem it is pointed out that we are not only dealing with a differentiation in a vertical sense, but also with a qualitative differentiation. Choice of a higher level-type of secondary education also means a high probability of entering into another type of work: more intellectual instead of practical, more concerned with abstract thinking instead of concrete thinking.

A sample was constructed, consisting of children who had just entered schools on the three main levels. Starting with the two most extreme levels: grammar schools (indicated by M) and trade schools (indicated by L), the score on all items was examined for level-differences by means of the statistical test of WILCOXON (MANN-WITNEY). Table 4.1 presents the findings: low figures indicating preference by the lower level pupils, high figures preference by the grammar school pupils. Statistical significance is indicated by means of asterisks (* $P < .05$; ** $P < .01$; *** $P < .001$). On examining the items with significant differences it was found that trade school boys and girls preferred items from the TH-, AV-, VB-, AA- and HA-scales, whereas grammar school pupils showed a very marked preference for TN- and LG-items (Cf table 4.2).

Because of the argument that differences in level may be of a qualitative as well as of a quantitative kind, it was decided to make two educational

level scales, one consisting of items preferred by lower-level pupils, the other by higher-level pupils. The scales are indicated by the letters L and M. It was found that on the great majority of all items concerned, pupils on the intermediate educational level (U) have an average score that is in between those of the upper and lower level pupils.

For scale construction only those items were selected that reached a very pronounced level of significance and also were chosen sufficiently frequently to be of practical value. Moreover, only those items were chosen on which the score of the intermediate level pupils was indeed in between those of the upper and lower level ones. Table 4.3 summarizes these items.

Preliminary norms were obtained on both scales and a cross-check made on a new sample of children. It was found that the amount of overlap did not increase considerably. A first attempt was made to establish the concurrent validity of the scales, by testing the power to distinguish between the three main level groups. It was found that both scales differentiated very significantly between upper and lower level candidates and that the L-scale was best suited to differentiate between upper and intermediate, the M-scale to differentiate between lower and intermediate level.

Apart from this attempt at level-scale construction, another was made to construct a weighted scale, especially for selection for grammar schools. The procedure outlined by KUDER ⁽⁸¹⁾ was followed. Results are reported in the sixth section of this chapter.

Correlations between interest-scale-scores and intelligence prove to be generally low, cf table 4.10.

Chapter V. The psychometric properties of the Vocational Interest Test

In this chapter data are reported on the homogeneity of test items, inter-correlations and reliability of scales.

Coefficients of homogeneity were calculated in accordance with the formula of LONG-LOEVINGER ⁽⁹²⁾. The procedure is outlined in the second section and data on item-scale homogeneity are presented in table 5.1. Only a few items did not prove to be satisfactory in their scale; in most cases coefficients are gratifying. The HA-scale proved to be an exception and so an attempt was made to revise this scale, by eliminating a number of items and including some from other scales. It was possible to make up a new scale, consisting of 16 items, of fairly good homogeneity (cf table 5.2).

Intercorrelations between the scales reported in table 5.3 are very similar to those found by IRLE, except in the cases of the HA and AD-scales. This is thought to indicate that the attitude towards both commercial and clerical work in this age group is less well established than that towards other fields of activity.

Scale-reliability was calculated as split-halves (page 1 vs page 2) and as KUDER-RICHARDSON, formula 21 reliability. Contrary to expectation the

former method gave rather lower values than the second one. As a check a number of odd-even correlations were computed, which, however, gave no consistent amelioration of the split-halves values. Only in the case of the L-scale, where the page 1 vs page 2 methods admittedly can be criticized because of the rather uneven distribution of the scale items, the odd-even methods gave very significantly higher values. Data on reliability are reported in table 5.4.

Chapter VI. Establishing norms

Norms on the Vocational Interest Test were established on a random sample of sixth grade children. The size of the sample was 1018, consisting of 531 boys and 487 girls. The method by which the sample was obtained is described in detail and an extensive check on its randomness is reported.

Norms for boys and girls separately were established in accordance with the procedure chosen by IRLE, i.e. percentile norms were calculated and projected upon a normalized scale. Figures 6.1 and 6.2 present the norms for girls and boys.

A simple way for testing significance between scale-scores was sought. After considering several possibilities, a method was chosen which, with a slight sacrifice of accuracy, enables the test user to check the significance on the profile-sheet by means of a ruler.

Also the problem of chance scores was discussed. An entirely random response to all test items yields an average scale-score of 18, with a standard deviation of 3.67. Chance values are indicated upon the figures 6.1 and 6.2 by a shadowing. It is thought that the influence of giving a considerable number of chance responses will be that ever more scale-scores will be near the shadowed values. This can be checked by counting the number of times the scale-score is on the same side of the median value as the shadowing. If this number is nine or more, the possibility of considerable chance influence should be taken into serious consideration.

Chapter VII. Using the test for establishing interest patterns

As a preliminary to the second part of the study some theoretical problems are discussed. First of all comes that of test validity. The recent developments concerning the concept of validity are mentioned and an investigation is made into the importance of the different kinds of validity for our subjects. Then the problem of an operational definition of interest is elaborated. It is pointed out that there is no direct relation between the definition given in chapter I and the operational definition of interest in terms of test scores. However, this relation can be established in an indirect way by means of the "bootstraps"-effect discovered by CRONBACH and MEEHL⁽⁴⁰⁾. The last section discusses the use of interest patterns especially for construct and concurrent validation.

Chapter VIII. Sex differences in the Vocational Interest Test

Sex differences in interest are well-known. Everyday observation shows that they exist already at the age studied in our investigation, though few details have been established in a scientific way. However, the hypothesis can be made that the main sex characteristic is rather independent of age. On this hypothesis a comparison can be made of sex differences as shown by the Vocational Interest Test and e.g. those shown by the KUDER-Preference Record. This test was chosen because it closely resembles our test and also because IRLE does not give information on this subject. It was possible to compare the two tests on seven scales; each time the sex difference in the median scores was in the same direction, so that the agreement is very significant, as checked by means of the sign-test.

Next the preference-pattern within sex was compared for boys and girls. It was found that the two scales preferred most by boys are TH (mechanical) and AA (agricultural). Next comes TN (science and engineering). Those preferred least are AD (clerical) and SO (social and educational). The other scales take an intermediate position. Girls show a very marked preference for the SO-scale, AV (artistic) comes second. Very little interest is shown for both the TN and TH-scales. Table 8.2 shows the average scores of boys (*jongens*) en girls (*meisjes*). Lastly the sex differences are investigated scale by scale and expressed by means of *phi* (point-fourfold) coefficients, a plus-value indicating that the scale is preferred most by boys. The possibilities of making a masculinity-femininity-scale are discussed, which is thought to be quite feasible.

Chapter IX. Choice of secondary education and interest-testscore

The material used for establishing norms also yielded information on the destination of children after leaving elementary school. The characteristic patterns of the different sub-groups are analyzed in this chapter. The interest pattern was characterized by means of a network, drawn from the test norms. The reference-points in this network are the raw scores just below the nine successive deciles. A cumulative frequency distribution of the scores on each scale is given per subgroup, using the nine reference-points. In each table the position of the median individual per scale is indicated by italics. In the last line of each table the number of individuals per scale, scoring at the ninth decile or higher, is given.

Successive tables in this chapter give the interest pattern of girls (9.1) and boys (9.2) going to grammar school, girls (9.3) and boys (9.4) going to a school at the intermediate level, girls choosing a trade school (9.5) or a school for continued elementary education (9.6) or leaving school after the sixth grade, because of super-annuation (9.7). It is found that in the group characterized by table 9.6 there are many girls with mechanical interests. In the current Dutch educational system they lack the op-

portunity to gratify their interests. Table 9.8 gives the interest pattern of boys going to a trade school, 9.9 those choosing continued elementary education, 9.10 those leaving school and 9.11 those choosing a career in agriculture. Lastly 9.12 and 9.13 present the interest patterns of girls and boys who plan to take an extra year of elementary education before applying for admission to a grammar school.

Chapter X. Interest test scores and choice of specific branches within the trade schools for boys

In the same way as in the preceding chapter interest patterns are presented for a number of branches within the trade schools for boys. Most data refer to boys that have just entered this type of school. In a few instances, however, also some second-year pupils have been included in order to increase the size of the sample. Tables refer to: machinists (10.1), instrument- and toolmakers (10.2), electricians (10.3), plumbers (10.4). A comparison of the interest patterns of these four groups shows a clear relationship. An arrangement from plumbers over machinists and electricians to toolmakers shows the characteristic features of this "job family" in an ever increasing way. Next the characteristic pattern of painters (table 10.5), paving bricklayers (10.6), carpenters (10.7) and cabinetmakers (10.8) are shown. These four branches also belong to one family. This time the typical line of arrangement is from paving bricklayer over carpenter and cabinetmaker to painter. The interest pattern of bakers, mainly pastry-bakers, is shown in table 10.9, those of composers and printers in 10.10 and 10.11. The similarity of the last two patterns is very marked. This is contrary to the generally held opinion that printers have a pronounced mechanical interest and are rather different from composers. Our figures show that both groups have the type of interest usually associated with that of composers only. Data published by IRLE on an older age-group of printers' apprentices show a similar interest pattern.

Several of our test groups correspond to groups examined by IRLE in Germany, with the only difference that on the average his subjects are about three years older than ours. Thus this study also affords an opportunity to investigate some age-influences upon interest. Table 10.12 presents the average scores on nine scales for seven occupations. Averages refer to percentile scores, which were computed on the basis of IRLE's norms-tables. In each group of two lines the upper one refers to the German, the lower to the Dutch subjects. On close inspection three consistent differences can be found: the older subjects generally have higher scores on the TN (science and engineering) and AD (clerical) scales and lower ones on the SO (social work and education) scales. Otherwise there is a remarkable similarity between the agegroups in the same branches, showing that age influences have little effect upon scores in this test.

Chapter XI. The predictive validity of the test

Data presented in the preceding chapters indicate a considerable concurrent validity of the Vocational Interest Test. Also some indications of construct validity can be found. This chapter now gives the results of an attempt at establishing the predictive validity of the testscores. The attempt was undertaken for one specific group, the grammar school pupils. Some problems arising in this connection are discussed, of which an important one is that of the criterion. For lack of a good achievement-test for this type of education, the promotion at the end of the first school-year was chosen as a criterion. This, however, though probably the best one available, is not very good, mainly because of very large variations in the degree of severity in the way pupils are judged at the end of the year. Differences between schools are shown by the extreme values, in one school 95% of the pupils were promoted, in another only about 60%. The problem of the relation between test score and criterion is also discussed, which may be linear or curvilinear. It is argued that the former supposition is the more probable one.

Over 500 grammar school pupils were examined. The total material first yielded an opportunity to check the findings presented in chapter IX. The tables 11.1 and 11.2 corresponding to 9.1 and 9.2 show that there is indeed a good confirmation of the expected patterns. Then pupils who are promoted to the second grade are compared with those who fail in their first year. Tables 11.3 and 11.4 present the patterns of success and failure for boys, tables 11.5 and 11.6 those for girls. Compared with the general populations there is little difference between those who are successful and those who fail, but on close inspection one finds that successful pupils show the typical grammar school interest pattern in a more marked way. The picture becomes even more pronounced, if the patterns of boys (11.7) and girls (11.8) who pass a rather severe selection, are studied. Differences are significant for boys on the TN and M-scales (high) and the L-scale (low). For girls a low score on the HA-scale is very characteristic. The other differences are not significant, though they are suggestive because of their consistency: high TN and M-scores, low L- and VB-scores.

Results show that the Vocational Interest Test indeed has some predictive validity. A study is planned to investigate the predictive validity over a longer range.

Attempts were made to investigate the validity of a weighted scale, as described in chapter IV. Two rather small samples of boys and girls, matched for intelligence, but different in school success, were studied. In this very restricted material only the low HA-score for girls and a high weighted scale score for both sexes gave a significant differentiation. However, a check made on larger material, not matched for intelligence, did not confirm the predictive powers of the weighted scale. In this in-

stance the original M-scale proved to be much more effective in differentiating between successful and unsuccessful pupils. In this case also two groups of pupils who had repeated the first grade and in their second attempt either had succeeded or failed to reach the second grade, were included in the study. It was found that these groups both on the M- and on the weighted scale had mean scores in between those of pupils for the first time in the first grade and succeeding or failing in their first attempt. It is thought that this indicates some formative influence of the educational environment, which is of little importance, when compared with the interests of the general population, but which becomes important when intra-school-group differences are studied.

Chapter XII. Personal data and interest test score

An investigation was made of the relation between scale scores and age when leaving the sixth grade, size of community, occupational level of father, occupational branch of father and type of denomination of the elementary school the child visits.

The slight age differences can easily be contributed to differences in the level of intelligence, the older pupils are those who have repeated classes. On the average they are less intelligent than the younger pupils. The size of the community seems to have little influence, the only differences found are between rural population and inhabitants of cities with over 500,000 people. These differences can be described mainly by stating, that the city-children show a pattern, which is somewhat nearer to that of the grammar school pupils. The few differences corresponding to occupational level of the father are again interpreted in terms of differences in intelligence. This does not hold good for the differences associated with the occupational branch of the father. These differences, too, generally are of little practical importance.

The problem of the possible relation between interest pattern and denomination is important, because differences have been found between the vocational preference of adherents of different religions at the university entrance level. At first it seemed that such differences, too, could be found in our material, especially because the Roman Catholic girls showed an interest pattern very dissimilar to that of other girls. However, it was found that this phenomenon was due to an intervening factor, related to differences in destination after leaving elementary school. When the samples were equalized for this influence, denominational differences practically disappeared. It is concluded that the differences in vocational preference at the university entrance level seem to be due to factors that are operating mainly during adolescence.