

DIE INVLOED VAN TOETSVOLGORDE
OP TOETSPRESTASIE AS
'N BATTERY VAN TOETSE TOEGEDIEN WORD

deur

DAVID JOHANNES GOUWS.

'n Verhandeling aangebied vir die graad

MAGISTER SCIENTIAE

aan die Potchefstroomse Universiteit

vir Christelike Hoër Onderwys.

Potchefstroom.

Januarie 1953

VOORWOORD	iv
HOOFSTUK I: ALGEMENE INLEIDING EN PROBLEEMSTELLING...	1
A. INLEIDING	1
B. PROBLEEMSTELLING	
1. Algemeen	4
2. Probleemstelling vir Eksperimentele Doelindes	5
HOOFSTUK II: OORSIG OOR DIE LITERATUUR	
BETREFFENDE TOETSVOLGORDE EN DIE FAKTORE WAT AFWYKENDE TOETSPRESTATIES KAN VEROORSAAK	7
A: DIE INVLOED VAN TOETSVOLGORDE OP TOETSPRESTATIE	7
B. DIE WISSELWERKING TUSSEN EMOSIONELE GESTELD- HEID EN TOETSPRESTATIES VAN PROEFPERSONE	
1. Algemeen	11
2. Die Differentiële Uitwerking van Emosionele Faktore op sekere Sielkundige Toetse	13
C. SAMEVATTENDE OORSIG	16
HOOFSTUK III: DIE NAVORSING OOR DIE VRAAGSTUK VAN DIE INVLOED VAN TOETSVOLGORDE OP TOETSPRESTATIES	18
A. INLEIDING	18
B. DOELSTELLING	18
C. DIE SIELKUNDIGE MEETINSTRUMENTE WAT IN DIE ONDERSOEK GEBRUIK IS.	
1. Die Persoonlike Vraelys	
a. Algemeen	19
b. Die vasstelling en meet van emosionele onstabiliteit	20
c. Die keuse van 'n geskikte vraelys	22
d. Bepanning van die nuwe vraelys	23
e. Betroubaarheid	29
f. Geldigheid	31
2. Die AG-toets	50
3. Die A/8-toets	52

4. Die M-toets	53
D. DIE PROEFPERSONE	
1. Algemeen	54
2. Eerste Eksperiment	56
3. Tweede Eksperiment	59
E. DIE PROEFNEMERS	60
F. DIE UITVOERING VAN DIE TOETSPROGRAM	
1. Algemeen	61
2. Eerste Eksperiment	62
3. Tweede Eksperiment	63
4. Hertoetsing met die Parallele Toetse	64
G. STATISTIESE TEGNIEKE WAT VIR DIE VERWERKING VAN GECEWENS GEBRUIK IS	
1. Korrelasie	64
2. Statistiese toetse vir beduidendheid	65
H. DIE EKSPERIMENTELE RESULTATE	
1. Gelykwaardigheid van die A- en B-groepe ten opsigte van toetsprestasies	66
2. Die Subjektiewe Beoordeling van die Toetssituasie	70
a. Die moeilikheid van die toetse	71
b. Die gewildste toets	75
c. Eie prestasie in die toetse	77
d. Vermoeidheid	79
e. Motivering	80
3. Die Invloed van Toetsvolgorde	81
a. Toetsprestasies in die eerste eksperiment	81
b. Toetsprestasies in die tweede eksperiment	83
c. Proefpersone met 'n hoë vraelyspunt	84
d. Die invloed van toetsvolgorde op gemiddelde vraelyspunt	86
e. Toetsvolgorde en instruksieduur	87
I. SAMEVATTING VAN DIE EKSPERIMENTELE RESULTATE ..	89

HOOFSTUK IV: 'N KRITIESE-SAMEVATTENDE BESKOUING

VAN DIE ONDERSOEK	93
A. ALGEMEEN	
1. Inleiding	93
2. Kritiese Samevatting van die Gangbare beskouings oor toetsvolgorde en die invloed van emosionele faktore op toetsprestasies	93
B. DIE RESULTATE VAN HIERDIE ONDERSOEK	
1. Hoofbevindings	94
a. Die invloed van toetsvolgorde op toetsprestasies	94
b. Die rol wat emosionele faktore speel	94
2. Bepelkings van die Eksperimentele OndersoeK	95
a. Die bepaalde toetse wat gebruik is	95
b. Die bepaalde proefpersone wat gebruik is	95
c. Die statistiese benadering van die vraagstuk	96
3. Die Persoonlike Vraelys	97
C. FINALE SAMEVATTING EN BESLUIT	99
1. Samevatting	99
2. Besluit	100
BIBLIOGRAFIE	101
BYLAE A.	

VOORWOORD.

Hierdie studie is gedurende 1951, terwyl die skrywer personeellid van die Wetenskaplike en Nywerheidsnavorsingsraad se Nasionale Instituut vir Personeelnavorsing (N.I.P.N.) was, uitgevoer. Die vraagstuk van die invloed van toetsvolgorde op toetsprestasie het in die loop van die daaglikse toetspraktijk van die N.I.P.N. opgeduik, maar nadat dit eenmaal duidelik geformuleer is, is beseef dat dit heeltemal algemeen is, en byna in alle psigometriese werk kan voorkom. Die vraagstuk is trouens so breed dat net enkele beperkte aspekte daarvan in hierdie studie ondersoek kon word.

Die indeling van die verhandeling is vierledig: Eerstens word aangetoon hoe die onderhawige vraagstuk onder verskillende omstandighede in die toetspraktijk ontstaan, en hoe dit vir die doeleindes van eksperimentele ondersoek geformuleer is, tweedens word 'n saaklike oorsig oor die verwante literatuur gegee, derdens word verslag gedoen oor die eksperimentele ondersoek en bevindinge en laastens volg 'n kritiese bespreking van, en samevattende beskouing oor, die vraagstuk van toetsvolgorde.

By die aanbieding van hierdie verhandeling is dit vir die skrywer 'n genoë om sy dank en erkentlikheid uit te spreek teenoor die volgende persone en liggame wat met hulle hulp en belangstelling hierdie ondersoek moontlik gemaak het:

Die direkteur en personeellede van die N.I.P.N;
Die direkteur van die Transvaalse Onderwysdepartement;
Mnr. J. P. Lehman, inspekteur van onderwys, Johannesburg;
Mnr. E. Botes, hoof van die Helpmekeer-hoërseunskool,
Johannesburg;
Dr. T. A. du Plessis, hoof van die Rossmore-hoërskool,
Johannesburg;
Mnr. J. J. van der Walt, hoof van die Fakkel-hoërskool,
Johannesburg;
Die Buro vir Opvoedkundige en Maatskaplike Navorsing,
Pretoria.

Die skrywer is ook besondere dank verskuldig aan sy promotor, prof. J. M. Hattingh, vir sy opbouende kritiek en gewaardeerde hulp, veral in die moeilike en belangrike fase van die opskrifstel van die studie, en aan sy vrou, Hester Gouws, vir haar hulp gedurende al die verskillende fases van die ondersoek.

HOOFSTUK I.

ALGEMENE INLEIDING EN PROBLEEMSTELLING

A. INLEIDING.

Hierdie studie is onderneem om die vraagstuk van die invloed wat toetsvolgorde op toetsprestasie uitoefen, te ondersoek. Om aan te toon hoe die vraagstuk in die praktyk ontstaan, word omstandighede in die Nasionale Instituut vir Personeelnavorsing (N.I.P.N.) by wyse van illustrasie beskrywe. In die N.I.P.N. is tot onlangs van 'n battery van vier toetse, naamlik vir Verstandelike Wakkerheid (AG), Rekenkundige Vermoë (A/S), Werktuigkundige Begrip (B) en Abstrakte Redeneervermoë (M) in 'n verskeidenheid van keuringsprojekte gebruik gemaak, onder andere vir die keuring van rekrute vir die leer en die vlootmag. Hoewel die volgorde waarin hierdie toetse aangebied is gewoonlik konstant gehou is, naamlik AG-B-A/S-M, het praktiese omstandighede soms afwyking van die neergelegde volgorde genoodsaak. 'n Voorbeeld hiervan is die geval waar 'n aantal proefpersone dieselfde toetsbattery moet aflê, en een of meer van hulle laat is vir die toetsproses. Hulle arriveer dan wanneer die res van die groep al een of meer van die toetse afgelê het. In sulke gevalle was dit gebruikelik om die laatkommers vir die oorblywende toetse van die battery by die groep te laat aansluit, en om die ontbrekende toetse daarna op hulle toe te pas.

Op grond van die veronderstelling dat sekere toetse beter deug vir aanvangstoetse as ander, het verskillende proefnemers die geldigheid van afwyking van 'n neergelegde toetsvolgorde betwyfel. Die moontlikheid is genoem dat

een toetsvolgorde beter of swakker toetsresultate sou lewer as 'n ander, wat individue sou bevoor- of benadeel, namate hulle die toetse in verskillende volgordes sou aflê. Verskillende verklarings waarom sekere toetsvolgordes as waarskynlik geskik, en ander as waarskynlik ongeskik beskou is, is aangebied. Daar is byvoorbeeld beweer dat die proefpersoon- en veral die senuagtige proefpersoon- aan die begin van die toetssessie nog nie voldoende aan die nuwe omstandighede aangepas is nie, en dat sekere toetse minder gevoelig is vir hierdie faktor, en ook meer geskik om aanpassing te bewerkstellig, as ander. Toetse met ingewikkelde instruksies sou waarskynlik nie goed deug vir aanvangstoetse nie, omdat proefpersone aan die begin van die toetssessie terughoudend is en nie geredelik vrae stel nie.

Van die vier toetse in die genoemde battery was dit veral toets M wat as ongeskik vir 'n aanvangstoets beskou is. Die instruksies is ingewikkeld, en die toetstaak is vir die meeste proefpersone nuut en vreemd. Aangesien die toets vir 'n verstandelik-meerderwaardige groep bedoel is, is die moeilikheidsvlak van die meeste items hoog. Toetse AG en A/8 is vanweë hulle skoolse aard weer as geskik vir aanvangstoetse beskou. Die aard van die toetstaak is aan die proefpersone bekend (sillogistiese redenasie, reken, ens.), en die instruksies vir hierdie toetse is ook betreklik eenvoudig.

Oorwegings soos die voorgaande was hoofsaaklik op ongekontroleerde waarneming van proefpersone in die toetssituasie gebaseer. In 'n paar gevalle is persone wat die M-toets eerste gedoen, en baie lae punte behaal het, oorgetoets, en het hulle aansienlike verbeterings in hulle punte/

hulle punte bewerkstellig. Hoe dikwels dit die geval was, en wat die bydrae van hertoetsing met dieselfde toets daartoe was, was egter onbekend.

Die verskynsel dat 'n groot aantal proefpersone of kandidate dieselfde battery van toetse moet aflê, word al hoe meer algemeen in psigometriese werk. Gedurende die Tweede Wêreldoorlog het die Amerikaanse Lugmag hulle lugbemanning met behulp van 'n battery bestaande uit ongeveer veertien gedrukte en ses apparaattoetse gekeur en geklassifiseer (8, pp. 71-109). Dieselfde verskynsel, hoewel op kleiner skaal, het ook die keuringsmetodes van die Statebondslande se militêre organisasies gekenmerk (38, p. 43; 2; pp. 14-15).

Klaarblyklik is dit in tyd van nasionale nood, wanneer groot getalle rekrute binne 'n kort tydsbestek getoets moet word, van belang om te weet of van 'n neergelegde volgorde van toetsaanbieding afgewyk kan word sonder om die doeltreffendheid van die toetsprosedure te verminder. Maar ook in die vredestrydse sielkundige navorsing en praktyk word al hoe meer van toetsbatterye in plaas van enkele toetse gebruik gemaak. In Thurstone se studie oor die primêre verstandelike vermoëns is van 'n battery van ses-en-vyftig verskillende toetse gebruik gemaak (37, p.10). Hoewel die volgorde van die toetse in Thurstone se ondersoek konstant gehou is, is die moontlikheid dat die bepaalde toetsvolgorde wat gebruik is 'n eie bydrae tot die navorsingsresultate kon gelewer het, nie uitgesluit nie. Op kliniese gebied word daartans baie waarde aan die toetsprofiel of patroon van toetsprestasies van proefpersone geheg. Wechsler het 'n diagnostiese skema uitgewerk wat grootliks op die toetsprofiel berus (39, pp. 146-167). Freeman maak die volgende bewering oor die gebruik van sielkundige toetse:

„In more/ ...

"In more recent years, however, without denying the usefulness and value of these indexes of relative status, (soos verstandsonderdom, intelligensiekoeffisient) + increasing emphasis has been placed upon 'patterns' of performance as clinical aids to psychological diagnosis and counseling" (11, p.28). In hierdie verband is dit van belang om te weet of 'n toetsprofiel wat met 'n bepaalde toetsvolgorde verkry is, dieselfde sou wees indien van 'n ander volgorde van dieselfde toetse gebruik gemaak is.

Uit die voorgaande blyk dus dat die vraagstuk van toetsvolgorde heeltemal algemeen is - dit kan oral waar daar in psigometriese werk van toetsbatterye gebruik gemaak word, opduik. Die vraagstuk kan in die praktyk ontwyk word deur op 'n bepaalde aanbiedingsvolgorde te besluit en daarmee vol te hou, maar die onbuigsaamheid van so 'n prosedure is dikwels ongerieflik en onprakties. Dit is dus wenslik dat vasgestel word of toetsvolgorde wel 'n bepalende invloed op toetsprestasies uitoefen, en waarom dit gebeur. Dit is ook van akademiese belang vir die sielkunde om, indien vasgestel kan word dat toetsvolgorde toetsprestasies bepalend beïnvloed, 'n verklaring daarvoor in terme van bekende sielkundige begrippe te probeer vind.

B. PROBLEEMSTELLING.

1. Algemeen.

Die vraagstuk van die invloed wat deur toetsvolgorde op toetsprestasies uitgeoefen word, is om verskillende redes baie breed. Dit betrek benewens die kognitiewe ook die affektiewe en wilsaspekte van die persoonlikheid, 'n Eksperimentele ondersoek van die vraagstuk word bemoeilik deurdat daar soveel verskillende soorte sielkundige toetse is, byvoorbeeld individuele en groep-toetse, geskrewe en handelingstoetse, aanlegs- en ~~Deur skrywer ingevoeg~~ Prestasietoetse.

prestasietoetse, ensovoorts. 'n Omvattende ondersoek sou die kenmerke van rangorde en wisselwerking van toetse van dieselfde soort, sowel as van verskillende soorte, moes vasstel. Die aantal moontlike verskillende rangskikkings in volgorde neem vinnig toe as die aantal toetse in die eksperimentele battery vermeerder word. Om die eksperimentele ondersoek binne hanteerbare afmetings te hou was dit dus noodsaaklik dat die veld van ondersoek streng beperk word. In 'n voorlopige ondersoek soos hierdie is dit ook wenslik dat die aantal veranderlikes in die eksperiment, (wat uiteindelik met mekaar in verband gebring moet word), so laag moontlik gehou word, anders word die vertolking van eksperimentele resultate buitengewoon moeilik.

Soos uit die voorafgaande bespreking in die Inleiding (pp. 1-4), en die latere bespreking van die literatuur oor toetspraktyk (pp. 7-17) blyk, word daar veral aan steurende emosionele reaksies en gesteldhede as verklaringsbeginsels vir afwykende toetsprestasies gedink. Dit was dus nie net nodig om vas te stel of verskillende volgordes van dieselfde toetse tot verskillende gemiddelde toetsprestasies aanleiding gee nie, maar ook om te bepaal wat die rol van sekere emosionele faktore daarby is.

2. Probleemstelling vir eksperimentele Doeleindes.

In die eksperimentele ondersoek moes

- (a) Vasgestel word of, en tot watter mate, die twee verskillende aanbiedingsvolgordes van twee potlood-erpapier toetse die toetsprestasies van proefpersone beïnvloed.
- (b) As vasgestel kon word dat toetsvolgorde toetsprestasie beduidend beïnvloed, bepaal word wat die bydrae van sekere emosionele faktore tot die waargenome verskille

in toetsprestasies/.

Die verband tussen aantal toetse en aantal rangskikkings is sodanig dat n toetse in $n!$ verskillende volgordes gerangskik kan word.

in toetsprestaties is.

Voordat die eksperimentele ondersoek in besonderhede beskryf kan word, moet eers aangetoon word tot watter mate die probleem van toetsvolgorde al deur ander navorsers teengekom en ondersoek is, watter gegewens oor die uitwerking van toetsvolgorde op toetsprestasiereeds beskikbaar is, en wat die gangbare beskouings oor die verband tussen emosionele gesteldheid en toetsprestasiere is.

HOOFSTUK II.

OORSIG OOR DIE LITERATUUR BETREFFENDE TOETSVOLGORDE EN DIE FAKTORE WAT AFWYKENDE TOETSPRESTATIES KAN VEROOKSAAK.

A. DIE INVLOED VAN TOETSVOLGORDE OP TOETSPRESTATIE.

Die vraagstuk van toetsvolgorde het reeds op 'n vroeë stadium in die ontwikkeling van die psigometrie, naamlik in die werk van Binet en Simon, opgeduik. Volgens Burt (4, p.9) het Binet, wat sy intelligensieskaal s6 opgestel het dat daar vir elke ouderdomsvlak (kronologiese jaar) 'n aantal toetse gerangskik in volgorde van stygende moeilikheid was, terwille van praktiese oorwegings aanbeveel dat hierdie volgorde waar nodig gewysig moes word. Sy instruksies/

'Al die reeds beskikbare uitgawes van „Psychological Abstracts (28), dit wil s6 jaargange 1 tot 25, en 'n aantal handboeke oor toetspraktyk (4, 5, 11, 26, 29, 32, 34, 36, 39) wat waarskynlik as verteenwoordigend van die veld beskou kan word, is vir die nasporing van bronne geraadpleeg. Verder is die personeellede van drie van die grootste toetsontwikkelingskorporasies in die Verenigde State, naamlik „Educational Testing Service“ (ETS), „The Psychological Corporation“ en „Science Research Associates“ deur middel van 'n omsendbrief geraadpleeg in verband met die opsporing van oorspronklike werk oor die vraagstuk van toetsvolgorde.

Sy instruksies was dat 'n toetssessie moes begin by die toets wat die geskikste vir die kind se kronologiese ouderdom geag is. As dit nodig geblyk het, moes terugwerk word na die toetse vir laer ouderdomme, totdat die kind in ten minste vyf agtereenvolgende toetse geslaag het, daarna moes weer in die voorwaartse rigting gewerk word, totdat die kind in ten minste vyf agtereenvolgende toetse gedruip het. Hoewel Burt die beginsel onderskryf dat die toetse in volgorde van toenemende moeilikheid toegedien moet word, waarsku hy teen 'n slaafse navolging daarvan (4,p.9) Van die sewe oorwegings vir die bepaling van toetsvolgorde deur hom genoem (4,pp. 9-10), skyn vier vir die onderhawige ondersoek van belang te wees, nl. (1) Die eerste toetse moet as skokbrekers dien, hulle doel is om die kind se senuagtigheid te laat bedaar en om sy belangstelling op te wek (2) Sodra die toetsers 'n aanduiding het van tussen watter ouderdomme hy moet toets, moet hy met die makliker en korter toetse begin. Enigiets wat die kind moontlik kan ontmoedig of vermoed moet nie te vroeg in die sessie gegee word nie. (3) Nietemin moet toetse wat wakkerheid en inspanning vereis nie tot op die laaste uitgestel word nie. Sulke toetse moet gegee word "When the child has warmed up to the process - having overcome his initial shyness, not yet jaded or bored, pleased perhaps by his success over the simpler tests, not yet disheartened by his failure in the harder" (4,p.10). (4) Toetse wat apparaat vereis, of liggaamlike bewegings van die kind, kan vir laaste gehou word. As daar egter tekens is dat die kind as gevolg van emosionele uitputting of vermoedienis begin faal, kan hulle ook vroeër gebruik word. Volgens Burt was dit vir Simon gebruiklik om die praktiese toetse oor die hele sessie te versprei, om die kind sodoende gedurig deur middel van afwisseling te verfris (4,p.10).

Hier is dus 'n voorbeeld van hoe kliniese oorwegings daartoe kan lei dat van 'n neergelegde toetsvolgorde afgewyk word. Hoewel geleidelike aanpassing aan die toetssituasie, die invloed van senuagtigheid op toetsprestasies, ens. geldige kliniese oorwegings vir die wysigings van toetsvolgorde is, is hier geen aanduidings van hoe verskillende volgordes die toetsprestasies en die geldigheid en toepaslikheid van die norme beïnvloed nie. Deur van die standaard-aanbiedingsvolgorde af te wyk, word die toetsprestasies op 'n ongekontroleerde wyse en tot 'n onbekende mate beïnvloed. In dié verband is dit van belang dat Terman en Merrill die volgende instruksies met betrekking tot toetsvolgorde vir die 1937 - hersiening van die Stanford-Binet-intelligensietoets gee: „The tests of each year group should be given in the order in which they appear in the manual and record booklet. The order is one which gives a minimum of monotony and also takes account of the effect of one test upon another.” (34,p.56). Aangaande die praktyk om al die toetse van 'n bepaalde tipe agtereenvolgens toe te dien, sê hulle: „Serial testing ... is not permissible if the test score is to be interpreted in terms of the established norms. ... The method in question is indefensible if for no other reason than that it changes the difficulty of the test items by an unknown amount. Perhaps a mental set for repeating digits makes this test easier, but a special experiment would be necessary to determine whether and to what extent this is true” (34,pp. 56-57). Hier is daar dus 'n uitgesproke erkenning van die problematiese aard van toetsvolgorde.

Cattell beskou die vraagstuk van toetsvolgorde as belangrik. Sy aanbevelings vir die vasstelling van toetsvolgorde lui onder andere soos volg:

„Generally speaking, it is best to start with performance tests, which evoke zest, and help to dissolve any feeling of strain or strangeness felt by the subject. Also one should start with easy tests, or at least tests which do not permit the subject to realize how far he is failing” (5, p. 267). Aangesien Cattell geen eksperimentele getuënis vir sy standpunt noem nie, moet aangeneem word dat dit bloot op kliniese mening berus. Dit is derhalwe opmerklik dat sy aanbeveling ten opsigte van handelingstoetse die teenoorgestelde van Burt s'n is. Die erkenning van die feit dat die subjektiewe beoordeling van die prestasie deur die proefpersoon 'n belangriker oorweging as die objektiewe moeilikheidsvlak van 'n toets is, is betekenisvol. Volgens Thorndike (36, p. 261) word die volgorde waarin die afsonderlike toetse van 'n battery toegedien word, gewoonlik deur oorwegings van gericflikheid bepaal, S6 is dit gewoonlik gerieflik om al die toetse wat dieselfde soort bekwaamheid vereis saam te groepeer. Hy erken egter ook die belangrikheid van afwisseling, veral vir die behoud van belangstelling, en vir 'n volgehoue poging deur die proefpersoon.

Super beskou die belangstelling van die proefpersoon en sy bereidwilligheid om sy bes te doen as die belangrikste oorwegings vir die vasstelling van toetsvolgorde. Hy noem die volgende algemene beginsels om hierdie doelstellings te bereik: „The first test in the series should be something of a buffer, one on which the examinee can warm up, get some self-assurance and develop some interest. For this reason it should not be too hard, should be relatively impersonal (i.e. neither an intelligence nor a personality test) and objective, and should have 'face validity' or seem pertinent to the reason for taking tests. Next should come a test or tests with long and difficult directions, difficult content, or other characteristics which

make/.....

make desirable an alert mind, ability to concentrate, and willingness to apply oneself" (32, p. 78).

Dit blyk uit die aangehaalde beskouings dat toetsvolgorde as belangrik beskou word, en dat daar 'n algemene neiging is om aan sekere sielkundige faktore soos aanpassing by die toetssituasie, aan- of ontmoediging, belangstelling, ens. as die bepalende oorwegings vir die vasstelling van toetsvolgorde te dink. Hierdie oortuigings berus deurgaans op kliniese waarneming en interpretasie, en geen eksperimentele gegewens word ter staving van die bewerings genoem nie. Dit was dan ook nodig om, voordat tot 'n eksperimentele ondersoek oorgegaan is, die emosionele reaksies wat met sielkundige toetsing gepaard gaan, en hulle wederkerige invloed op toetsprestasies, in meer besonderhede in die literatuur te ondersoek.

B. DIE WISSELWERKING TUSSEN EMOSIONELE GESTELDHEID EN TOETSPRESTASIES VAN PROEFPERSONE.

1. Algemeen.

Hoewel daar reeds sedert die tyd van Binet en Simon baie oor die noodsaaklikheid van die daarstelling van rapport tussen proefnemer en proefpersoon gepraat en geskryf is, bewys die skaarste aan eksperimentele gegewens op dié terrein dat die belangrikheid van emosionele faktore as medebepalers van toetsprestasies dikwels onderskat of heeltemal buite rekening gelaat word. Burt het in dié verband gewaarsku: „Too often it is forgotten that every test of intelligence is also a test of emotion. A neglect of the emotional aspect of an examination of intelligence may send many a normal child to the special school as mental defective" (4, p. 16). In dieselfde gees merk Cattell op dat met verontagting van emosionele faktore selfs nie van intelligensie 'n

betroubare maat moontlik is nie (5, p. 268).

Proefpersone beskou emosionele faktore as van groot belang, en die uitgesprokene beweer dikwels dat hulle ten tyde van die toets te opgewonde, of gespanne, of bevrees gevoel het, om hulle beste prestasies te kon gelewer het. Dit is des te meer opmerklik dat daar so min akkurate kennis oor die uitwerking van verskillende emosionele toestande op toetsprestasies beskikbaar is,

Van 'n belangwekkende ondersoek van Yager word die volgende opsomming gegee: „Forty boys, aged 10 - 12, were tested under normal conditions and under emotional tension induced by a threat to their integrity and evidenced by physiological changes. There were thirty gains and ten losses in score on the Otis test under emotional tension. These changes were positively related to emotional stability (40, p. 464). Dit sou vir die onderhawige ondersoek van groot waarde gewees het as 'n aanduiding gegee is van hoe 'n maat van die emosionele stabiliteit van die proefpersone bekom is.

Die ervaring van personeellede van die N.I.P.N. wat die Wechsler-Bellevue op 'normale' proefpersone sowel as op kliniese gevalle toegepas het, skyn daarop te dui dat senuagtige persone meer verward en ontsteld raak as hulle toetse met tydsgrense moet aflê, as wanneer hulle toetse van onbeperkte duur moet doen. Cattell (5, p.5) maak 'n soortgelyke opmerking. In 'n bespreking van foute wat by die toediening van sielkundige toetse gemaak word, noem Ligon (21, p. 392) as 'n belangrike faktor die verskynsel dat proefpersone vanweë hulle ervaring met skoolksamens waarin hulle verwag word om alles wat gevra word te kan beantwoord, 'n gevoel van minderwaardigheid kry as hulle in 'n sielkundige toets vrae teenkom wat hulle nie kan beantwoord nie. Sy oortuiging

is dat/.....

is dat sulke proefpersone die stryd dan te gou gewonne gee, en dat hulle toetsprestasies dus nie getroue weergawes van hulle bekwaamhede is nie.

Die gegewens dusver betref sielkundige toetsing in die algemeen. Onderliggend aan hierdie ondersoek is egter die gedagte dat sekere toetse meer gevoelig is vir die nadelige uitwerking van emosionele faktore, en daarom minder geskik vir aanvangstoetse, as ander is. In die volgende afdeling word hierdie aspek van die wisselwerking tussen emosionele gesteldheid en toetsprestasie bespreek.

2. Die differensiele uitwerking van emosionele faktore op sekere sielkundige toetse.

Dat sekere kliniese sindrome kenmerkende simptome het wat met behulp van diagnostiese toetse geïsoleer en herken behoort te kan word, word deur sommige ondersoekers as aksiomaties aanvaar. Die neiging van pogings soos dié van Wechsler (39), Rapaport (29) en andere, is juis om 'n meer objektiewe diagnostiese skema as die gangbare daar te stel, deur van die differensiele uitwerking van bepaalde kliniese toestande op toetsprestasies gebruik te maak.

Rapaport vind dan ook by die toepassing van die Wechsler-Bellevueskaal dat angs by proefpersone dikwels tot visuele disorganisasie lei (29, p. 259), en dat angsneurotici se prestasies op die voorwerpsamestellings-toets (object assembly) beduidend benadeel word (29, p. 261). Prestasies op die syferomvangstoets (digit span) word ook opmerklik deur angs benadeel (29, p. 193). Volgens Rapaport is die rekenkundetoets in die Wechsler-Bellevueskaal egter nie baie gevoelig vir angs nie (29, pp. 207-209).

Bewerings soos die voorgaande impliseer dat die proefnemer om een of ander rede die aanname maak dat die proefpersoon wat aan angs ly, beter op die bepaalde toetse sou kon gepresteer het as hy nie deur angs benadeel

was nie./.....

was nie. Die geldigheid van só 'n aanname hang van die inter-korrelasies van die toetse in die battery wat gebruik word af. Objektief kan egter bewys word dat sekere toetse meer gevoelig is vir bepaalde emosionele gesteldhede as ander deur die gemiddelde toetsprestasies van 'n groep proefpersone by wie 'n bepaalde toestand aanwesig is, met dié van 'n onuitgesoekte groep te vergelyk. Só 'n ondersoek is deur Halstead (17) uitgevoer. Hy het by die ontleding van 700 neurotiese (militêre) proefpersone se prestasies in Raven se „Progressive Matrices 1938” - toets gevind dat die neurotiese groep statisties-beduidend laer prestasies as 'n normale kontrolegroep behaal het (17,p.202). Hierdie bevinding kon volgens een of albei van die volgende beginsels verklaar word: (a) die differensiese uitwerking van 'n neurotiese toestand op die betrokke toets, en (b) die laer algemene potensiese prestasievlak van die groep neurotiese proefpersone in vergelyking met die bevolking.

Aangesien in psigometriese werk geen absolute, maar slegs 'n relatiewe maat van prestasie moontlik is, kon die geldigheid van die bogenoemde verklaringsbeginsels slegs getoets word deur toetse van verskillende vorm en inhoud, maar wat veronderstel is om dieselfde vermoë, bv. algemene intelligensie, te meet, aan die proefpersone te gee, en vas te stel of daar opmerkbare verskille in die gemiddelde prestasies op die verskillende toetse is.

Halstead (17) het benewens die „Progressive Matrices 1938” - toets ook die Shipley-woordeskattoets op die 700 neurotiese proefpersone toegepas, en gevind dat hulle prestasies op hierdie toets, hoewel nog laer as dié van die bevolking, nie soveel deur hulle toestand benadeel

is as wat met die „Progressive Matrices 1938”-toets die geval was nie (17,p. 211). Dit skyn dus of die „Progressive Matrices 1938”-toets gevoelig is vir emosionele wanaangepastheid. Halstead maak inverband met sy ondersoek die volgende belangwekkende opmerking: „The results also confirmed observations made during test sessions that subjects who scored a high grade on the Matrix Test tended to be less disturbed by their neuroses on written tests of various kinds” (17,p. 202).

In 'n rapport oor die toetsprestasies van oudsoldate in beskutte arbeid toon Goldstein (12) aan dat die gemiddelde prestasies van die proefpersone op die „Progressive Matrices 1938” - toets in verhouding met die norme aansienlik laer is as die gemiddelde prestasies op die Wechsler-Bellevueskaal. Haar gevolgtrekking is: „There are indications that the discrepancies between the Matrix and Wechsler-Bellevue ratings are greater for persons in whom symptoms of anxiety are present” (12,p.13).

Volgens Vernon en Parry (38,p.239) het rekrute wat as gevolg van twyfelagtige emosionele stabiliteit deur leerpsigiaters afgekeur is, as 'n groep besonder sleg gevaar in die sg. „Squares”-toets, 'n toets van ruimtelike oordeelsvermoë.

Hoewel dit uit hierdie oorsig van die literatuur blyk dat bevindings betreffende die differensiele uitwerking van bepaalde emosionele gesteldhede nog van 'n voorlopige aard is, lyk dit baie waarskynlik dat prestasie in toetse met 'n sterk visuele komponent, nadelig deur toestande soos angs, angsneurose, ens. beïnvloed word, terwyl prestasie in toetse van 'n meer verbale aard waarskynlik minder ernstig deur die genoemde toestande geraak word.

C. SAMEVATTENDE OORSIG.

Die belangrikste bevindings waartoe die oorsig oor die literatuur gelei het, is die volgende:

- (1) Gesaghebbendes op die gebied van sielkundige toetsing beskou die probleem van toetsvolgorde as belangrik.
- (2) Op grond van toetservaring word verskillende soorte toetsvolgordes aanbeveel. Hoewel die besonderhede van voorgestelde volordes soms bots (Burt vs. Cattell, 1938, p. 10) is daar 'n algemene erkenning van die belangrikheid van die proefpersoon se emosionele orientasie ten opsigte van die toetssituasie.
- (3) Daar is getuienis van 'n algemene aard dat sekere emosionele gesteldhede soos senuagtigheid, gespannenheid, minderwaardigheidsgevoelens, ens., prestasies in sielkundige toetse nadelig beïnvloed.
- (4) Dit skyn asof emosionele gesteldhede soos senuagtigheid, gespannenheid, minderwaardigheidsgevoelens, ens., 'n nadeliger uitwerking op die toetsprestasies van emosioneel-wanaangepaste persone het as wat met „normale” persone die geval is.
- (5) Dit kan as betreklik seker beskou word dat sekere soorte toetse gevoeliger is vir die uitwerking van afwykende emosionele gesteldhede as ander. Dit skyn veral met toetse wat ruimtelike oordeelsvermoë vereis die geval te wees.

'n Verdere gevolgtrekking waartoe hierdie oorsig oor die literatuur lei is, dat daar 'n algemene gebrek aan eksperimentele gegewens oor die vraagstuk van die wisselwerking tussen emosionele gesteldheid en toetsprestasie is. Sonder twyfel is hierdie toestand ook die gevolg van die komplekse aard van die veld. Op hierdie stadium van ontwikkeling van sielkunde as

vakwetenskap is daar skynbaar nog geen bevredigende, .
omvattende teorie van emosie nie, en daar is skynbaar
geen betroubare tegnieke vir die vasstel en meet van
sulke verskynsels soos emosionele onstabiliteit beskik-
baar nie. Hierdie toestand van sake lei enersyds tot
versigtigheid in die aanvaarding en vertolking van
beskikbare gegewens, en andersyds tot die besef dat
navorsing op hierdie gebied baie noodsaaklik is. In
die volgende hoofstuk word aangetoon hoe 'n eksperimen-
tele ondersoek van die vraagstuk van toetsvolgorde
uitgevoer is.

HOOFSTUK III.

EIE NAVORSING OOR DIE VRAAGSTUK VAN DIE INVLOED VAN TOETSVOLGORDE OP TOETSPRESTASIE.

A. INLEIDING

Die besondere vorm wat hierdie eksperimentele ondersoek van die vraagstuk van toetsvolgorde aangeneem het, is hoofsaaklik deur die volgende omstandighede bepaal: Ten eerste het die vraagstuk in die loop van die daaglikse toetspraktyk van die N.I.P.N. ontstaan. Dit het die keuse van toetse vir die eksperimentele battery grootliks beïnvloed. Ten tweede is die ondersoek deurgaans só beplan dat die gegewens met behulp van verskillende statistiese tegnieke behandel sou kon word. Veral was dit vir die toepassing van statistiese toetse vir beduidendheid wenslik dat betreklike groot groepe proefpersone gebruik moes word. Omdat hulle betreklik maklik in groot getalle getoets kon word, en omdat daar ook fasiliteite vir die kies van gelykwaardige groepe beskikbaar was, is skoolkinders deurgaans as proefpersone gebruik. Derdens het dit in hierdie ondersoek veral om die vasstelling van die bydrae van emosionele faktore tot waargenome afwykende toetsprestasies gegaan. Aangesien geen geskikte instrument vir die meet van emosionele wanaangepasheid van skoolkinders gevind kon word nie, is 'n poging aangewend om 'n vraelys wat 'n betroubare en geldige maat van emosionele wanaangepasheid sou gee, te ontwikkel.

B. DOELSTELLING

Dit is nou moontlik om die doelstelling van hierdie bepaalde eksperimentele ondersoek in presieser bewoording te stel, nl.

(a) Om vas te stel of, en tot watter mate, die twee

verskillende/.....

verskillende aanbiedingsvolgordes van¹ (i) toetse AG en M; (ii) toetse A/8 en M die toetsprestasies van gelykwaardige groepe skoolkinders beïnvloed;

- (b) Om, as vasgestel kon word dat toetsvolgorde die toetsprestasies beduidend beïnvloed, te bepaal wat die bydrae van sekere emosionele faktore, soos met behulp van 'n aanpassingsvraelys vasgestel en gemeet, tot die waargenome verskille is.

C. SIELKUNDIGE MEETINSTRUMENTE WAT IN DIE ONDERSOEK GEBRUIK IS.

1. Die Persoonlike Vraelys.²

- (a) Algemeen. Die hipotese wat in hierdie studie ondersoek word is dat indien verskillende volgordes van dieselfde toetse die toetsprestasies verskillend sou beïnvloed, die emosionele toestand van die proefpersoon, en die emosionele reaksies op bepaalde toetsvolgordes belangrike veroorsakende faktore sou wees. Meer in die besonder is vermoed dat proefpersone wat geneig is om senuagtig, mindorwaardig, terneergedruk, mismoedig, ens. te voel, d.w.s. persone met negatiewe selfgevoelens, swakker sou presteer as hulle aan die begin van die toets-sessie al dadelik 'n baie moeilike en geheel-nuwe toets moes doen. Die subjektiewe reaksie op so 'n ondervinding sou vermoedelik 'n nadelige uitwerking op prestasies in daardie toets, en moontlik ook op prestasies in die daaropvolgende toetse hê. Dit was dus nodig om (i) 'n aanduiding van die algemene emosionele gestelheid, die kenmerkende gevoelston en houdinge van die proefpersone, sowel as (ii) 'n aanduiding van die proefpersone se subjektiewe ervaring in, en beoordeling van, die toetssituasies te verkry.

(b) Die/.....

-
1. Vir 'n bespreking van hierdie toetse sien pp. 50-54
2. Omdat hierdie 'n nuwe instrument is, word dit breedvoeriger bespreek.

(b) Die vasstelling en meet van emosionele onstablieit.

'n Logiese voorvereiste vir die herkenning en meting van persoonlikheidsienskappe is dat daar eenduidige, definieerbare kategorieë van persoonlikheidsgedraginge bestaan, en ook norme of maatstawwe om die mate van besit van, of gebrek aan, bepaalde kenmerke mee aan te dui. Geen van beide is op die huidige stadium van ontwikkeling van die persoonlikheidsielkunde met enige mate van finaliteit beskikbaar nie - daar is 'n veelheid van stelsels van kliniese kategorieë, en geen betroubare instrument, of verteenwoordigende norme, om die posisie van die individu ten opsigte van die bevolking mee aan te dui nie. Wexler beweer dan ook: "The problem of personality evaluation is still one of the most perplexing on the psychological research agenda" (39, p. 129).

Subjektiewe beoordeling van persoonlikheidskenmerke van proefpersone, met of sonder die hulp van gestruktureerde skattingskale, is nie net moeilik en duur, in terme van gespesialiseerde personeel en tyd vereis nie, dis ook nog nie bevredigend ten opsigte van die betroubaarheid en geldigheid van die resultate nie, tensy die kenmerke wat beoordeel moet word, en die manier waarop dit gedoen word, betreklik eenvoudig is nie (38, pp. 152-164). Aangesien in hierdie ondersoek van skoolkinders as proefpersone gebruik gemaak is, en byna 700 proefpersone die toetse afgeleë het was dit al net om die tydsfaktor buite die kwessie om van enige onderhoud of ander individuele metode vir die beoordeling van persoonlikheidskenmerke gebruik te maak. 'n Vraelys, waarmee 'n aanduiding van die algemene emosionele gesteldheid, sowel as van die subjektiewe reaksie op die toetssituasie verkry kan word, en wat onder gekontroleerde omstandighede onmiddellik na voltooiing van die toetse beantwoord kon word, was blykbaar die enigste metode wat goed by die res van die eksperimentele prosedure ingepas kon word, hoewel dit

goed/...

goed beseef is dat die vraelys as diagnostiese en meet-instrument waarskynlik aan nog ernstiger gebreke mank gaan as die onderhoud.

Die besware teen die vraelysmetode is legio. Vernon se o.a.: „One of the difficulties of personality testing is that, when subjects realize the object of the test. (or if they misinterpret the object) they will always tend to respond so as to put themselves in the most favourable light” (38,p. 254). Eenwens die feit dat 'n proefpersoon die funksie van 'n vraelys op 'n bepaalde manier vertolk, en sy reaksies dienoreenkomstig kan wysig, is daar egter nog 'n fundamentele probleem, nl. of proefpersone wat werklik eerlik probeer wees met die beantwoording van 'n vraelys, voldoende selfkennis en oordeelsvermoë besit en 'n getroue weergawe van hulle persoonlike kenmerke te kan gee. Rapaport beweer dat die direkte vraelys op sy beste 'n getroue weergawe van die proefpersoon se bewuste houdinge en gedagtes kan gee, en dat die moontlikheid om selfs dit te kry, gering is (29,p. 396). Aangesien die proefpersone in hierdie ondersoek skoolkinders was is dit van belang dat Symonds die mening toegedaan is dat bewuste oneerlikheid selde kinders se vraelysantwoorde beïnvloed. Hy sê o.a. „The writer has long been of the impression from casual observation that pupils in school will answer questionnaires rather truthfully. One reason for this is that the pupils confuse these questionnaires with tests, and with the general school attitude of trying to do their best on a test they seriously try to give truthful answers on questionnaires concerning even their most personal affairs (33,p. 154).

'n Ander belangrike beswaar teen die vraelysmetode is van semantiese aard - kan die vrae werklik duidelik

en/.....

en verstaanbaar gestel word, en het 'n vraag dieselfde betekenis vir die proefnemer en die proefpersoon, en van proefpersoon tot proefpersoon? Gesien saam met ander besware, wat onder die hoofde „betroubaarheid” en „geldigheid” bespreek word, blyk dit dus dat die vraelys uiteraard aan baie gebreke mank gaan. - Die grootskaalse gebruik van die vraelysmetode is dan ook deur verskeie gesaghebbende navorsers gekritiseer, o.a. deur Cattell: „... as a serious approach to discovering the structure of personality it (the questionnaire)¹ represents the nadir of scientific inventiveness and subtlety” (6,p. 341). Hy kwalifiseer egter ook sy kritiek: „The abuses practiced by those who typically seize upon the questionnaire are no argument, however, against its proper use within the true limits of its validity” (6,p. 342). Een van die gepaste gebruike van die vraelys-volgens Cattell- is vir voorlopige ondersoek, wat later met behulp van meer objektiewe metodes herhaal kan word (6,p. 343).

Die beslissing om 'n vraelys in hierdie ondersoek te gebruik is gedoen met inagneming van hierdie besware. Groot versigtigheid in die aanvaarding en vertolking van vraelysgegevens moes aan die dag gelê word, en enige gevolgtrekkings wat gemaak word, moes as van 'n voorlopige aard beskou word.

(c) Die keuse van 'n geskikte vraelys. Na ondersoek van 'n aantal vraelyste wat naastenby die veld dek wat vir hierdie studie vereis is, o.a. van die „Minnesota Personality Scale”, die „Brown Personality Inventory for Children”, Bell se „Adjustment Inventory (Student form)”, die „Minnesota Multiphasic Personality Inventory”, die „Cowan Adolescent Adjustment Analyzer” en die „Bernreuter Personality Inventory” is tot die gevolgtrekking gekom dat/....
↑Deur skrywer ingevoeg.

dat nie een van die beskikbare instrumente volledig aan die vereistes van hierdie studie voldoen nie. In elk geval sou 'n vertaling nodig wees, en dit is as twyfelagtig beskou dat blote vertaling van 'n vraelys wat oorspronklik in 'n ander taal, en onder ander kulturele omstandighede opgestel is, net so 'n goeie instrument as die oorspronklike sou lewer. Die vraelyste wat ondersoek is was ook in ander opsigte onbevredigend. Die onderliggende standpunte en aannames van die opsteller(s) is óf nie duidelik gestel nie, óf was in die lig van die voorafgaande kritiek teen die vraelysmetode onaanneemlik. Die oorwegings wat die keuse van verskillende tipes of terreine van emosionele wanaanpassing, en die relatiewe verteenwoordiging daarvan in die vraelys beheer het, is selde duidelik gestel en geregverdig.

'n Belangrike funksie van die vraelys sou wees om 'n aanduiding van die subjektiewe reaksie op, en beoordeling van, die toetssituasie te verkry. Deur 'n nuwe vraelys op te stel kon vrae betreffende hierdie aspekte van die ondersoek ingesluit word, en kon die hele vraelys meer spesifiek aan die vereistes van hierdie ondersoek aangepas word.

(d). Bepanning van die nuwe vraelys. Die voorafgaande kritiek teen die gebruik en samestelling van vraelyste stel reeds die vereistes vir die nuwe vraelys. Vir die skrywer was die belangrikste vereiste dat daar duidelikhed moes wees oor watter terreine van emosionele wanaanpastheid of onstabiliteit ondersoek moes word, en dat hierdie terreine eweredig en verteenwoordigend deur die items in die vraelys gedek moes word. 'n Duidelike omskrywing en afbakening van die begrip „emosionele wanaanpastheid” of „emosionele onstabiliteit” is dus 'n voorvereiste. So iets is egter nog nie in bevredigende

vorm/.....

vorm beskikbaar nie. Aangaande die veelheid van beskouings oor emosionele aangepastheid sê Wexler die volgende: „In view of the frequent use of the phrase 'measure of personal adjustment', one might assume that personal adjustment is a clearly definable concept. Actually there exist a host of diverse, even mutually exclusive, definitions. If it had been necessary to settle on the precise nature of 'personal adjustment' before undertaking the measurement of its variable manifestation, we would even now be scratching fruitlessly in the boneyard of discarded phrases" (39, p. 136).

Die doel van hierdie ondersoek het daartoe gelei dat die begrip van wanaangepastheid vir die doeleindes van die samestelling van 'n vraelys baie breed gestel is. Alle emosionele afwykings wat toetsprestasies moontlik kon beïnvloed moes ingesluit word. Aangesien die vraelys nie bedoel was om die aard en samehang van emosionele steurnisse, of die moontlikheid om daarin selfstandige en herkenbare kliniese entiteite te vind, te ondersoek nie, is só 'n breë formulering as aanneemlik beskou.

Die ontwikkeling van empiriese wetenskappe vind plaas as 'n proses van wisselwerking tussen teoretiese formulering en feitelike waarneming, waarin die teorie die feitelike struktuur al beter benader. Dit sou dus onrealisties en nutteloos wees om op 'n bevredigende formulering van die verskynsel van wanaangepastheid te wag, voordat met navorsing begin word. Al eis wat aan 'n nuwe hipotese of ondersoek gestel kan word - benevens innerlike konsistensie - is dat dit bestaande kennis in ag sal neem. 'n Samevatting van die meeste ondersoek wat met behulp van vraelyste gedoen is, is in die werk van Cattell (6) beskikbaar. Deur faktor-analitiese metodes het by die massa bevindinge tot 'n paar fundamentele faktore of sindrome van emosionele wanaangepastheid/...

wanaangepastheid gereduseer. Hoewel - volgens Cattell self (6, p. 342) - hierdie konstruksies met voorbehoud aanvaar moet word, bied dit seker die beste samevatting van eksperimentele resultate wat op hierdie gebied beskikbaar is.

Deur die faktore wat reeds met 'n taamlike mate van sekerheid vasgestel is as basis te gebruik, kon die vraelys so verteenwoordigend moontlik van die verskillende wanaanpassings-brandpunte gemaak word. Die faktore wat as die vernaamste vir die doeleindes van hierdie ondersoek beskou is, is „Skaamheid“, „Algemene Emosionele Oorgevoeligheid“, „Swaarmoedige Verontrustheid“, „Algemene Neurotiese Wanaanpassing“, en „Senuagtige Angsigheid en Onstabiliteit“ (6, pp. 355-368). Omdat die vraelys vir so 'n homogene groep hoesel was, kon die verskillende vrae baie spesifiek gestel word. Vorige ondersoekers het gewoonlik gevind dat aanpassingsvraelyste, as dit op kinders toegepas word, lae betroubaarheid vertoon, derhalwe is 'n besondere poging aangewend om die vrae in hierdie vraelys so duidelik en ondubbelsinnig as moontlik te stel. Om 'n bevredigende aanduiding van die betroubaarheid van die vraelys te verkry is ook besluit om vir elke item wat vir insluiting in die vraelys gekies word 'n ooreenstemmende een op te stel wat so na as moontlik (volgens subjektiewe beoordeling) dieselfde vraag sou stel, maar in ander woorde. Die feit dat Cattell by elke faktor die items met 'n saturasie van 0.4 of hoër gegee het (6, p. 355), het die keuse van items en hulle ekwivalente baie vergemaklik. Sommige items, o.a. die betreffende stemmingsveranderlikheid, somatiesse simptome, ens. kom in dieselfde of byna dieselfde vorm by meer as een van die faktore voor. Sulke items is as besonder belangrik beskou. Die verskillende faktore hierbo genoem het naastenby gelyke verteenwoordiging in terme van aantal items in die vraelys gehad.

Aangesien die belangstelling in hierdie ondersoek op die toetssituasie saamgetrek is, is die items sover moontlik só opgestel dat dit betrekking op die skool-situasie het. Aangesien van die feit dat die skool-situasie vir kinders een van die belangrikste konfliktpunte is (9, pp. 59-60), is gevoel dat gebrekkige aanpassing op hierdie bepaalde gebied waarskynlik baie geredelik op die toetssituasie oorgedra sou word. Dit is as onwenslik beskou om items betreffende die verhoudings tussen ouer en kind, en tussen die twee ouers onderling, in die vraelys in te sluit.

'n Lys van 68 items, oor twee bladsye versprei sodat vir elk van die 39 items op die eerste bladsy daar 'n ooreenkomstige item op die tweede bladsy is, en van die „Ja - Nee - ?” - tipe, is vervolgens aan 'n aantal medestaflede in die N.I.P.N. voorgelê om veral ten opsigte van die taalaspek gekritiseer en verbeter te word. 'n Hele paar items is op die manier oorgeformuleer.

Dit is as noodsaaklik beskou dat hierdie voorlopige redaksie van die vraelys deur toepassing op 'n klein groepie proefpersone aangevul moes word. Die gewysigde vorm van die vraelys is derhalwe op 115 Std.6 - en Std.7 - seuns van 'n Afrikaanse hoërskool in Johannesburg, (3^e skool genoem) toegepas. Hierdie groepie proefpersone - wat origens nie in die ondersoek gebruik is nie - was waarskynlik verteenwoordigend van dié wat in die twee hoërskole waar die eintlike ondersoek uitgevoer is aangetref sou word, behalwe dat geen Std. 8 - seuns ingesluit is nie. Aangesien dit veral die doel was om moeilike bewoording en dubbelsinnigheid in die items te ontdek, sou hierdie verskil waarskynlik geen nadelige uitwerking/...

uitwerking nê nie. By hierdie toediening van die vraelys is die proefpersone aangemoedig om vrae te stel oor items wat hulle nie kon verstaan nie. Alle vrae wat gestel is, is deur die proefnemers aangeteken. 'n Verdere oogmerk van hierdie voorlopige toediening was om naastenby die totale toedieningsduur van die vraelys te bepaal, sodat die toetsprosedures by die ander twee skole noukeurig uitgewerk kon word.

Nadat die betroubaarheid van die vraelys vir die groep van 115 proefpersone bereken is,¹ is 'n itemontleding uitgevoer. Items wat dikwels deur die „?” alternatief beantwoord is, is as waarskynlik onduidelik beskou. Die item wat deur die grootste aantal proefpersone (nl. 30 uit 'n totaal van 115) met die „?” beantwoord is, het gelui: „Soek ander seuns graag jou geselskap?” Hierdie item is gewysig om te lees: „Dink jy dat ander seuns graag met jou wil maats wees?” Nog 'n aantal items is op grond van die itemontleding gewysig of uit die vraelys verwyder. Die vrae wat gedurende die voorlopige toepassing in skool C deur proefpersone gevra is, is ook by die hersiening van die vraelys in ag geneem. Sommige woorde wat skynbaar moeilik was, soos „lugkastele”, „kritiseer”, „hunker”, ens. is tog behou, omdat bespreking in die klas na afloop van die voorlopige toepassing, sowel as die resultate van die itemontleding, getoon het dat sulke onkunde net in geïsoleerde gevalle voorgekom het, en ook omdat die betrokke items beswaarlik gewysig kon word sonder om dit woordeliks identies met die ooreenstemmende items in die ander helfte van die toets te maak.

Die resultate van die itemontleding is ook gebruik

om/.....

1. Sien pp. 29 - 31.

om die aanvangsitoms in die finale vraelys naastenby in volgorde van dalende aanneemlikheid te plaas. As die veronderstelde „wanaangepaste” antwoord op 'n item deur 'n groot persentasie proefpersone gekies word, kleef daar vir hulle waarskynlik geen sosiale stigma aan daardie soort reaksie nie. Deur sulke items by of naby die begin van die vraelys te plaas, sal die natuurlike agterdog en vreemdheid van die proefpersone waarskynlik makliker oorkom word.

Na die finale redaksie en rangskikking van hierdie items, is die vrae wat spesifiek op die eksperimentele toetsituasie betrekking het, bygevoeg. Vier van hierdie vrae is op die buiteblad van die vraelys geplaas, en moes deur die invul van die betrokke toets se naam.

(d.w.s. AG, A/8 of M) beantwoord word. Vraag (a)¹ nl.

„Watter toets het jy die eerste gedoen?”, is net ingesluit om te verseker dat elke proefpersoon sou weet wat die name van die toetse is. Hierdie vraag is in die eksperimentele ondersoek met behulp van die proefnemer ingevul. Die eerste vier vrae op die eerste bladsy van die vraelys het ook spesifiek na die toetsituasie verwys en is in die „Ja - Nee -?” - vorm gestel. Die finale telling op die vraelys is verkry deur te tel hoeveel van die items die proefpersoon met die „wanaangepaste” alternatief beantwoord het. Die antwoorde op die vrae op die buiteblad, sowel as op vrae 1, 2 en 38, is nie by hierdie finale telling ingesluit nie. Vraag 38, nl.

„Hou jy baie van sport?”, is aan die begin van die tweede bladsy geplaas om die duplisering te vermoen en: origins gerusstellend te werk. Die hoogste moontlike telling was dus 70, en 'n hoë telling is veronderstel om aanduidend van 'n sterk mate van wanaangepastheid te wees. Die vraelys is die „Persoonlike Vraelys” genoem.¹

1. Sien Bylaag 1.

(e) Betroubaarheid/..

(e) Betroubaarheid. Een manier waarop die betroubaarheid van 'n toets gedefinieer kon word, is in terme van sy korrelasie met 'n parallelle vorm van dieselfde toets (16, pp. 13-14). In die ontwikkeling van die Persoonlike Vraelys is 'n bewuste poging aangewend om die eerste helfte gelykwaardig met die tweede te kry ten opsigte van vorm en inhoud van vrae. Indien die twee helftes as parallel beskou kan word sal met behulp van die korrelasie tussen die twee helftes 'n aanduiding van die betroubaarheid van die vraelys verkry kan word. Die betroubaarheid van die totale telling, gebaser op die som van die twee helftes, kon met behulp van die Spearman-Brown formule vir 'n dubbellengete-toets verkry word. Die voorwaarde vir die toepassing van hierdie formule is ook dat die twee helftes parallel moet wees (16, pp. 62-67). Hoewel bedoel om parallel te wees, voldoen die twee helftes egter nie aan die statistiese kriterium vir parallelle toetse nie, wat eis dat die gemiddeldes sowel as die variansies gelyk moet wees (16, pp. 13-14, pp. 173-174), omdat die verskille tussen die gemiddelde tellings op die twee helftes groter is as wat bloot op grond van steekproefvariasie verwag kan word. Gulliksen toon egter aan dat die Spearman-Brownformule goeie resultate lewer selfs wanneer geen besondere poging aangewend word om parallelle helftes te verkry nie (6, p. 67). Die gemiddeldes, standaardafwykings en interkorrelasies van die twee helftes, asook die gekorrigeerde korrelasiekoëffisiënt vir die skatting van vollengte betroubaarheid van die vraelys, soos vir die verskillende eksperimentele groepe bepaal, word in Tabel I aangegee, asook histogramme wat die verspreiding van vraelyspunte binne/.....

1. Hier soos ook in die volgende afdeling oor „Geldigheid” word die ontwikkeling van die ondersoek enigszins vooruitgehoop deurdat van resultate van die eksperimentele ondersoek van die invloed van toetsvolgorde op toetsprestasie gebruik gemaak word. 'n Volledige bespreking van die proefpersoon en die metode volg later in hierdie hoofstuk.

Gemiddeldes, standaardafwykings en Korrelasiekoeffisiënte van die twee helftes, en skattings van die vollengte betroubaarheid, van die Persoonlike Vraelys.

PROEFPERSONE		GEMIDDELDE TELLINGS				STANDAARDAFWYKINGS		r	\bar{k}
Groep	Aantal	1e bl.	2e bl.	1e bl.	2e bl.	1e bl.	2e bl.	(Vollengte betroubaarheid.)	
3e Skool	115	13.804	13.457	4.561	4.972	.747	.855		
1e Skool A-Groep	204	11.863	10.902	5.362	5.370	.817	.899		
1e Skool B-Groep	204	12.637	11.471	5.204	5.631	.806	.893		
2e Skool A-Groep	118	13.025	11.280	5.197	5.260	.768	.869		
2e Skool B-Groep	118	13.669	12.008	5.550	5.656	.808	.894		

TABEL I

binne die twee totale eksperimentele groepe aandui, in Figure 1 en 2 aangegee word.

Die eerste ry gegewens is op die voorlopige vorm van die vraelys gebaseer, en die res van die gegewens is vir die finale vorm van die Persoonlike Vraelys, soos dit in die eerste en die tweede eksperiment (onderskeidelik by die 1^e en die 2^e skool) gebruik is. Hieruit blyk dat die betroubaarheid van die vraelys as meetinstrument, selfs volgens die streng vereistes wat aan toetse vir kognitiewe vermoëns gestel word (15, p. 219), bevredigend is.

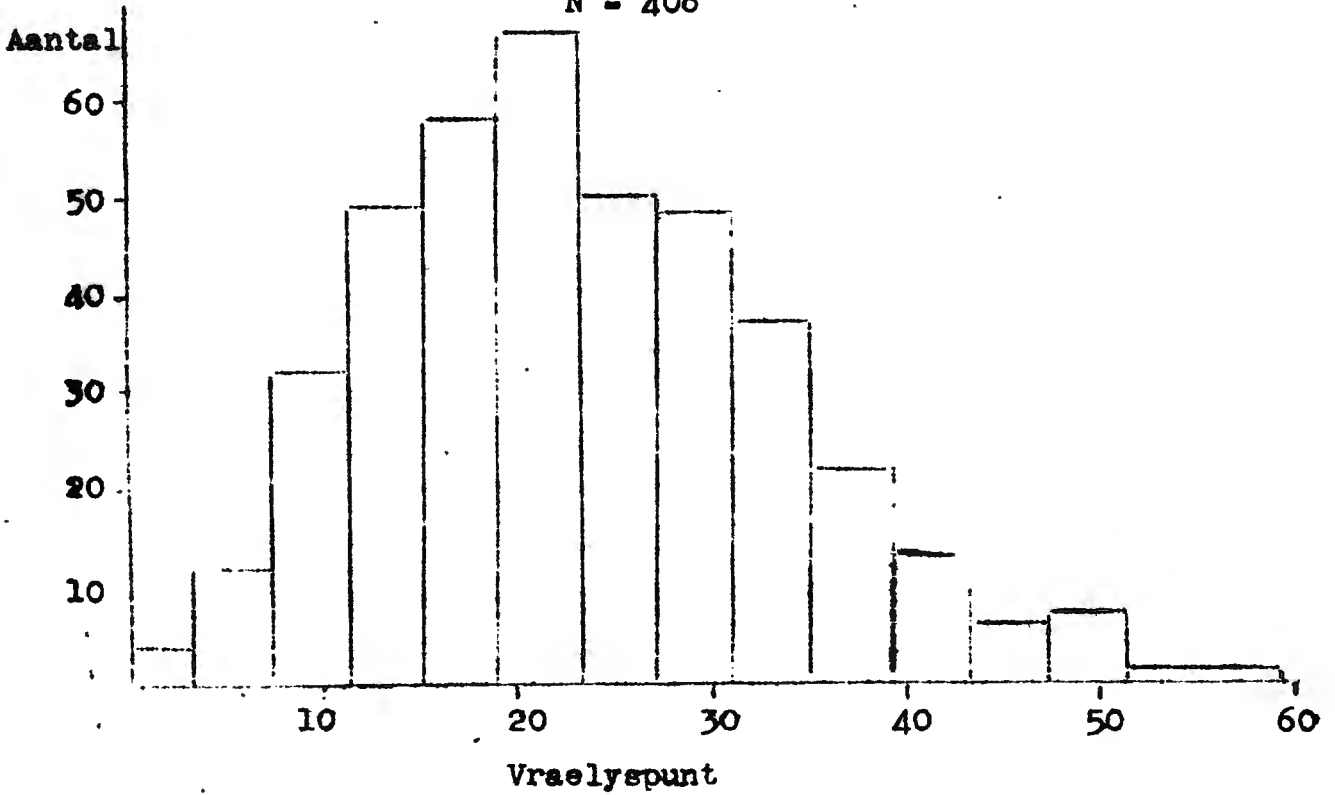
(f) Geldigheid. By die samestelling van die vraelys is gepoog om die volste gebruik van die beskikbare kennis in die literatuur te maak, en om die items in die vraelys so duidelik as moontlik te stel. Hoewel dit uit die voorgaande gegewens oor die betroubaarheid van die Persoonlike Vraelys blyk dat die vraelys konsekwent is in wat dit meet, moet nou nog aangetoon word dat dit wel meet wat dit veronderstel word om te meet, nl. algemene emosionele wanangepastheid, of watter naam ookal verkies word om die wye reeks afwykende emosionele gedraginge mee aan te dui. Soos Mosier (25, pp. 6 - 7) reeds aangetoon het is dit nie korrek om op grond van konsistensie en ooreenstemming van die vraelysinhoud met gesaghebbende mening sonder meer die geldigheid van 'n vraelys te aanvaar nie. Vir oortuigende aanduiding van die geldigheid van 'n sielkundige meetinstrument, ook van 'n aanpassingsvraelys, is eksterne validasie nodig.

Daar is twee hoofmetodes in swang vir die eksterne validasie van aanpassingsvraelyse. Volgens die eerste word bepaal watter mate van ooreenkoms daar tussen

vraelysresultate/.....

Verspreiding van Vraelyspunte van die Totale Groep Proefpersone in die Eerste Eksperiment

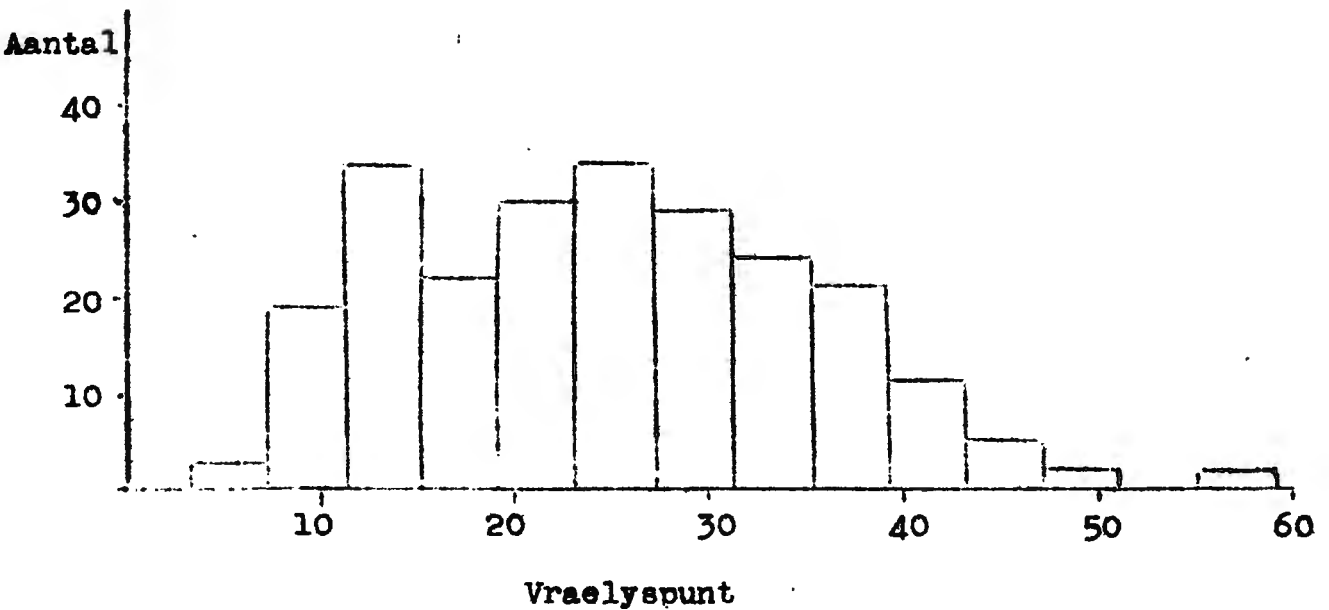
N = 408



Figuur 1.

Verspreiding van Vraelyspunte van die Totale Groep Proefpersone in die Tweede Eksperiment

N = 236



Figuur 2.

vraelysresultate en beoordelings van proefpersone deur deskundige personeel soos psigiaters of kliniese sielkundiges is. Op hierdie wyse word soms goeie geldigheidskoëffisiënte verkry. Conrad (aangehaal deur Vernon en Parry, 38, p. 257) het egter opgemerk dat die ooreenkoms wat wel gevind word, waarskynlik daaraan toegeskryf moet word dat psigiaters gedurende hulle onderhoude dieselfde soort vrae aan die proefpersone stel, as wat hulle in die vraelys moet beantwoord. Die fundamentele tekortkoming van hierdie validasiemetode is dat die geldigheid van beoordelings deur psigiaters of kliniese sielkundiges onkrities as geldig aanvaar word. Ten opsigte van psigiatriese keuring in die militêre situasie maak Wexler die volgende opmerking: „.... a comparison of test results with psychiatric diagnoses is merely a comparison of one prediction with another prediction. To speak of test validity on the basis of the outcome of such a comparison is to assume, a priori, the validity of the psychiatrists' judgements” (31, p. 138).

Die tweede manier waarop getrag word om die geldigheid van aanpassingsvraelyste aan te toon, is om die vraelys op 'n onuitgesoekte „normale” groep mense, en op 'n groep wat 'n neurotiese ineenstorting gehad het en gehospitaliseer is, toe te pas en dan aan te toon dat die vraelys die twee groepe doeltreffend onderskei. Vernon (38, p. 257) vestig die aandag daarop dat die „neurotiese” groep nie noodwendig op dieselfde wyse sou gereageer het as die vraelys vóór die ineenstorting op hulle toegepas is nie. Wexler se kritiek van hierdie validasiemetode is so raak en pittig dat dit, tesame met die voorbeeld wat hy ter staving van sy kritiek

noem/.....

noem, hier vollelig aangehaal word: „Nothing is so apt to produce delusions of grandeur in the constructor of personality inventories than the use of what may be termed the 'ex post facto' psychiatric criterion. This involves a technique by which a hospitalized population is contrasted with successful military personnel in a training center. The difficulty with this procedure is that patients in a neuropsychiatric ward have generally become so sensitized to clinical symptomatology that they no longer react in the way in which they would have reacted to the same instrument prior to hospitalization. Furthermore, the hospital wards generally contain a selected sample from the extreme of the deviated cases. --- It is no remarkable thing, then, that test score differentiation from a young, enthusiastic, and healthy recruit group can be obtained with considerable ease. As a single instance, one of the tests was administered to a neuropsychiatric ward population in a naval hospital. This group was found so far different from a random sample of qualified recruits drawn from a training center that 87% of the hospital group scored above an established critical score point whereas only 4% of the qualified recruit group was found within the same range of scores. Needless to say, the test was not found so efficient under more exacting circumstances in which adjusted and maladjusted individuals were intermingled in the same group and before deviates had been identified psychiatrically" (31, pp. 140-141).

'n Beter validasiemetode as die voorgaande is waarskynlik om die vraelyspunt met die mate van sukses

of/.....

of mislukking waarmee lewensituasies gehanteer word, te vergelyk. Omdat die manier waarop lewensituasies gehanteer word ook ander aspekte van die persoonlikheid as net emosionele aangepastheid betrek, bv. algemene intelligensie, kan 'n bepaalde mate of soort van wanaangepastheid een persoon gevoeliger tref as 'n ander, en moet die toerusting vir die lewe (die „erflike” potensialiteit), die lewensomstandighede, en die hantering en waardering van situasies dus tesame gesien word om die mate van aangepastheid te bepaal. In hierdie opsig is die standpunt van Horney onderskryf dat wanaangepastheid objektief in 'n wanverhouding tussen potensialiteit en werklike prestasie, tussen werklike lewensomstandigheid en subjektiewe waardering daarvan, waargeneem kan word (19, pp. 22-23). Die verdienste van hierdie beskouing oor wanaangepastheid is dat dit 'n moontlikheid vir objektiewe validasie van die vraelysmetode bied, mits 'n aanduiding van die proefpersoon se potensiële vermoëns, sowel as 'n aanduiding van sy werklike prestasie op 'n bepaalde belangrike gebied verkry kan word. In hierdie ondersoek was daar 'n maat van die proefpersoon se algemene intelligensie, wat algemeen as 'n aanduiding van die intellektuele toerusting vir skoolprestasie beskou word, in die intelligensiekoëffisiënte (I.K.'s) beskikbaar. Daar was ook, in die geval van die eerste eksperimentele groep (1^e skool) 'n maat van die proefpersone se werklike skoolprestasie, nl. uitslae in die halfjaarlikse skooleksamen, beskikbaar. Die geldigheid van die Persoonlike Vraelys kon dus bepaal word deur die vraelyspunte met die verskille tussen potensiële en werklike toetsprestasies te vergelyk. Soos reeds aangedui (p. 31), is dit in psigometriese werk meestal nodig om die validasieprosedure

krities te ondersoek om te sien of dit geldig is.

Die versweë aanname van die voorgestelde validasie-prosedure vir die Persoonlike Vraelys, is dat afwykende skoolprestasies uitsluitlik of grotendeels die gevolg van emosionele wanaangepastheid is. Hierdie aanname is slegs waarskynlik waar, omdat faktore soos sportbedrywighede, stokperdjies, gesondheid, verandering van skool, ens. ook die skoolprestasies bepalend kan beïnvloed. Hoewel die onsekerheid van die inherente geldigheid van die kriterium besef is, is die validasie tog uitgevoer, omdat die kriterium minstens betreklik onafhanklik en objektief was. By die vertolking van die verkreë geldigheidskoëffisiënte moet die onsekere aard van die kriterium egter in gedagte gehou word.

'n Faktor wat daartoe kan bydra dat 'n direkte verband tussen wanaangepastheid en afwykende skoolprestasies nie gevind sou word nie, is die feit dat I.K.'s met behulp van 'n reeks toetse vasgestel word, en dat wanaangepaste kinders „laer mag toets” as waartoe hulle in staat is, sodat die wanverhouding tussen potensiële en werklike prestasies nie duidelik word nie. Die voorgestelde prosedure is slegs uitvoerbaar as die maat van potensiële vermoë glad nie, of slegs in geringe mate, en die werklike prestasie wel tot 'n aansienlike mate, deur die mate van emosionele aangepastheid beïnvloed word.

Die oortuiging is gehuldig dat eksamenresultate, wat die prestasie oor 'n wye kennisgebied en oor 'n tydperk van etlike maande verteenwoordig, meer opmerklik deur wanaangepastheid beïnvloed sou word as wat die geval is met die prestasie in 'n enkele intelligensietoets, wat nie op spesifieke kennis berus nie, en wat in hierdie geval ook nie van die toetse insluit wat skynbaar/..

skynbaar baie gevoelig vir wanaangepastheid is nie. Intelligensiemeting behels meer die formele aspekte van die kind se intellektuele vermoë en ervaring, terwyl die skooleksamen gerig is op die meting van spesifieke kennis en vaardighede, wat oor 'n lang tydperk verkry is, en dus waarskynlik aan die kumulatiewe uitwerking van emosionele afwykings blootgestel was. Hoewel die geldigheid van hierdie opvatting betwisbaar is, word dit dikwels in besprekings van die opvoedkundige situasie implisiet aanvaar. Slegs deur aan te neem dat ons 'n aanduiding van 'n persoon se intellektuele vermoëns kan verkry wat minder deur emosionele steurnisse benadeel is as wat met eksamenresultate die geval kan bv. betekenisvol beweer word dat baie mislukings op skool en universiteit aan wanaangepaste persoonlikheidstrekke, wat niks met die intellektuele vermoëns van die persoon te doen het nie, te wyte is (9, pp. 59-60).

As aangeneem word dat die I.K. (soos vasgestel met behulp van die Suid-Afrikaanse Groepsverstandtoets) 'n geldige maat van die onderwysbaarheid (educability) is, en dat die vasstelling daarvan minder deur emosionele steurnisse beïnvloed word as wat met skooleksamens die geval is, moet nog aangetoon word dat skooleksamens betroubare en toepaslike (relevant) kriteria van skolastiese ontwikkeling is. Op grond van die werke van Ballard (1), en van Hartog en Rhodes (18), word algemeen aangeneem dat skooleksamens onbetroubare maatstawwe is. Oplekters het ook in 'n onlangse studie (27) aangetoon dat daar opmerkbare verskille van skool tot skool is in die standaard van werk binne dieselfde skoolstanderd.

In/.....

In hierdie geldigheidstudie is die eksamenresultate van verskillende klasse van dieselfde skoolstandord in dieselfde skool as vergelykbaar beskou, omdat al die klasse van 'n skoolstandord hulle onderrig in 'n bepaalde vak grotendeels van dieselfde onderwyser gekry het, en omdat hulle almal dieselfde eksamenvraestelle beantwoord het. Aangesien daar geen parallelle eksamen beskikbaar was nie, en 'n halveringskorrelasie van die eksamenprestasies ook nie moontlik was nie, is van 'n metode van betroubaarheidsberekening wat op die interkorrelasies van die punte in die verskillende vakke berus, gebruik gemaak. Die afleiding van die formule, wat die laer perk gee van die betroubaarheid van samegestelde punte wat op prestasies in afsonderlike toetse berus, is deur Sichel (30) gegee. Die interkorrelasies tussen die prestasies in die halfjaarlikse eksamen, 1951, in 6 vakke, nl. Afrikaans, Engels, Geskiedenis, Matesis, Rekenkunde en Natuurwetenskap word afsonderlik vir die drie skoolstandords en die A- en B-groepe van die eerste eksperiment (1^e skool) in Tabela 2 - 4 aangegee.¹ Die standaardafwykings van die punte vir elke vak, vir die drie skoolstandords en die A- en B-groepe, word in Tabel 5 aangegee.

1.- Behalwe vir een Std. 8-seun, was die halfjaarlikse eksamenpunte van al die proefpersone wat finaal in die eerste eksperiment gebruik is, beskikbaar.

Vak-interkorrelasies, Std. 6.

		Afrikaans	Engels	Geskiedenis	Natuurwetenskap	Rekenkunde	Matesis	
A-Groep N=85	Afrikaans		.828	.761	.219	.658	.738	
	Engels	.815		.735	.586	.629	.672	
	Geskiedenis	.782	.631		.664	.727	.752	
	Natuurwetenskap	.592	.472	.577		.546	.672	
	Rekenkunde	.688	.700	.721	.554		.727	
	Matesis	.736	.710	.765	.636	.715		
								B-Groep N=85

TABEL 2.

Vak-interkorrelasies, Std. 7.

		Afrikaans	Engels	Geskiedenis	Natuurwetenskap	Rekenkunde	Matesis	
A-Groep N=66	Afrikaans		.555	.473	.330	.197	.448	
	Engels	.607		.457	.202	.386	.370	
	Geskiedenis	.566	.559		.452	.378	.622	
	Natuurwetenskap	.497	.482	.591		.410	.657	
	Rekenkunde	.482	.314	.514	.523		.686	
	Matesis	.503	.324	.616	.622	.722		
								B-Groep N=66

TABEL 3

Vak-interkorrelasies, Std. 8.

	Afrikaans	Engels	Geskiedenis	Natuurwetenskap.	Rekenkunde	Matesis
A- Groep N=53.	Afrikaans	.418	.347	.202	.281	.188
	Engels	.585	.338	.333	.345	.257
	Geskiedenis	.391	.570	.560	.354	.353
	Natuurwetenskap.	.221	.354	.577	.328	.413
	Rekenkunde	.396	.308	.452	.342	.682
	Matesis	.125	.181	.394	.360	.664
						B- Groep N= 52

TABEL 4

Standaardafwykings van Eksamenpunte in elke vak, Stds. 6 - 8.

	A - GROEP			B - GROEP		
	Std.6 N=85	Std.7 N=66	Std.8 N=53	Std.6 N=85	Std.7 N=66	Std.8 N=52
Afrikaans	2.748	2.275	1.581	3.058	1.932	1.725
Engels	3.244	2.202	2.694	3.184	2.027	2.781
Geskiedenis	4.282	3.306	2.970	4.934	2.862	2.558
Natuurwetenskap	2.904	2.928	3.106	2.676	2.819	3.015
Rekenkunde	4.310	4.272	4.285	4.249	3.906	3.666
Matesis	3.323	3.695	3.505	3.246	3.691	3.609

TABEL 5.

Met hierdie gegewens was dit nou moontlik om die laer perk van betroubaarheid van 'n kriterium wat uit 'n samestelling van die genoemde 6 eksamenvakke bestaan, te bereken. Die berekende minimum-betroubaarheidskoëffisiënte van die saamgestelde kriterium vir die A- en B-groepe van Stds 6 - 8 word in Tabel 6 aangegee.

Minimum-betroubaarheid van 6-vekeksamen-
punte, Std. 6-8.

	A- GROEP			B- GROEP		
	Std. 6	Std.7	Std.8	Std.6	Std.7	Std.8
Minimum- betroubaar- heidskoëffi- siënt.	.943	.906	.869	.945	.869	.853
Aantal in groep.	85	66	53	85	66	52

TABEL 6

As in aanmerking geneem word dat 'n betroubaarheid van .90 of hoër vir verstandstoetse vereis word (15, p. 219), dan is hierdie koëffisiënte wat die minimum-betroubaarheid van die saamgestelde kriterium aandui, baie bevredigend.

Aangaande die geldigheid van skooleksamens as mate van skolastiese ontwikkeling en prestasie, blyk dit uit 'n oorweging van die manier waarop eksamenvraestelle opgestel word, dat dit toepaslik (relevant) kriteria van mate van beheersing van die stof in die leerplan behoort te wees, omdat so ver as moontlik getrag word om 'n verteenwoordigende steekproef van die onderwerpe wat behandel is, in die skooleksamen in te sluit.

Dit is ook opmerklik dat die matrikulasie-eksamen - waarin die verspreiding van talent onder die kandidate alreeds baie beperk is - aansienlike geldigheid vir die voorspelling van sukses in universiteitseksamens het, soos blyk uit 'n ontleding deur Biesheuvel (3) van eksamenprestasies in die Fakulteit Geneeskunde aan die Witwatersrandse Universiteit; en uit 'n vergelyking van matrikulasieuitslae met eerste- en tweedejaars-

B.Com.-uitslae van studente aan die Witwatersrandse Universiteit, wat korrelasieskoëffisiënte van die orde .70 opgelewer het.¹ Die gevolgtrekking dat skool-eksamens betroubare en geldige kriteria vir die meting van skolastiese ontwikkeling en prestasie is, skyn dus goed verantwoord te wees. Vervolgens kon met die voorgestelde prosedure vir die validasie van die Persoonlike Vraelys voortgegaan word.

Die taak om die Persoonlike Vraelys te valideer kon, volgens die opvatting van wanaangepastheid as 'n wanverhouding tussen potensiële en werklike prestasie, op twee maniere aangepak word. Albei metodes is geprobeer. Ten eerste is die korrelasies tussen die I.K.'s, die toetse AG en M, die Persoonlike Vraelys, en die gemiddelde 6-vak-eksamenpunte bereken, in die verwagting dat aangetoon sou kon word dat die vraelys 'n beduidende bydrae tot die meervoudige korrelasie tussen die voorspellers (toetse AG en M, I.K.'s) en die kriterium (eksamenpunte) lewer. Die betrokke korrelasiematrikse vir die A- en B-groepe in die drie skoolstanderds van die eerste eksperiment word in Tabelle 7 - 9 aangegee.

1. Ongepubliseerde materiaal in navorsingslêers, KI/51/7 van die N.I.P.N.

Korrelasiematrikse,
Std. 6.

	I.K.'s	AG	M	Vraelys	Kriterium (Eksamen- punte.)
A- Groep N=85	I.K.'s	.800 ^b	.376 ^b	.200	.766 ^b
	AG	.802 ^b	.481 ^b	-.176	.749 ^b
	M	.498 ^b	.526 ^b	-.089	.384 ^b
	Vraelys	-.221 ^a	-.263 ^a	.023	-.292 ^b
	Kriterium (Eksamen- punte.)	.778 ^b	.712 ^b	.462 ^b	-.248 ^a
					B- Groep N=85

TABEL 7

a: p <.05
b: p <.01

Korrelasiematrikse,
Std. 7

	I.K.'s	AG	M	Vraelys	Kriterium (Eksamen- punte.)
A- Groep N=66	I.K.'s	.524 ^b	.372 ^b	-.190	.491 ^b
	AG	.623 ^b	.618 ^b	-.007	.709 ^b
	M	.395 ^b	.478 ^b	.041	.429 ^b
	Vraelys	-.450 ^b	-.214	-.110	-.162
	Kriterium (Eksamen- punte.)	.533 ^b	.557 ^b	.265 ^a	-.207
					B- Groep N=66

TABEL 8

a: p <.05
b: p <.01

KORRELASIE-MATRIKSE, STD. 8.

	I.K.'s	AG	M	Vraelys	Kriterium (Eksamen- punte)
I.K.'s		.512 ^b	.504 ^b	.128	.454 ^b
AG	.478 ^b		.497 ^b	.002	.635 ^b
M	.407 ^b	.226		.162	.364 ^b
Vraelys	-.017	-.075	-.042		.114
Kriterium (Eksamen- punte.)	.380 ^b	.488 ^b	.141	-.145	

A- Groep N=55

B- Groep N=52

TABEL 9

a: $p < .05$
 b: $p < .01$

Die kritiese waardes van die korrelasiekoëffisiënt ooreenstemmend met die .01- en .05- waarskynlikheidsvlak is met behulp van Fisher en Yates se tabelle (10) vasgestel.

Die negatiewe korrelasie tussen vraelyspunte en eksamenprestasie was volgens verwagting, en nou moes vasgestel word of insluiting van die vraelys by die voorspellers tot beter voorspelling van die kriterium sou lei. Die meervoudige korrelasiekoëffisiënte tussen verskillende kombinasies van die voorspellers en die kriterium is vervolgens vir die A- en B-groepe van Std. 6 uitgewerk. Die koëffisiënte word in Tabel 10 aangegee.

Uit/.....

Meervoudige korrelasiekoëffisiënte,
Std. 6.

	A-Groep	B-groep
$R_{0.12}$.792	.799
$R_{0.123}$.794	.800
$R_{0.1234}$.796	.811

TABEL 10^a

^a Die R-indekse verwys soos volg: 0- Kriterium
1- I.K.'s
2- AG
3- M
4- Vraelys.

Uit Tabel 10 blyk dat die bydrae van die Persoonlike Vraelys tot die gesamentlike voorspelling van die kriterium gering is. Aangesien die geldigheidskoëffisiënte vir die Persoonlike Vraelys in Stds. 7 en 8 kleiner as dié in Std. 6 is (Tabelle 8 en 9) is dieselfde prosedure nie vir die twee standerds herhaal nie, In plaas daarvan is bepaal wat die meervoudige korrelasie van I.K.'s en vraelyspunte met die kriterium is in elk van die A- en B-groepe in die drie skoolstanderds. Hierdie koëffisiënte word in Tabel 11 aangegee. Om die bydrae van die vraelys te bepaal word die geldigheidskoëffisiënte van die I.K.'s alleen ook aangetoon.

Ook uit die gegewens in Tabel 11 blyk dit dat die Persoonlike Vraelys nie 'n opmerklieke bydrae tot die gesamentlike voorspelling van die kriterium lewer nie.

Die/.....

Meervoudige korrelasiekoëffisiënte

Std. 6 - 8.

	A- GROEP			B- GROEP.		
	Std. 6	Std. 7	Std. 8	Std. 6	Std. 7	Std. 8
$R_{0.1}$.778	.533	.380	.766	.491	.454
$R_{0.14}$.782	.534	.404	.779	.496	.457

TABEL 11^a

^aDie R-indekse verwys soos volg: 0- Kriterium
1- I.K.'s
4- Vraelys.

Die opvatting van wanaangepastheid as 'n wanverhouding tussen potensie en werklike prestasie lei egter ook tot 'n meer direkte metode waarvolgens die geldigheid van die vraelys vir die meet van wanaangepastheid ondersoek kan word. Dit is nl. om vas te stel wat die verband tussen vraelystelling en waargenome afwyking van die verwagte eksamenpunt is, waar die verwagte eksamenpunte deur middel van die regressie van die gemiddelde 6- vakeksamenpunte op I.K.'s verkry word. Soos uit Tabel 7 (p.43) blyk, is die korrelasie tussen I.K.'s en eksamenpunte van die orde .77-.78 vir die A- en B-groepe in Std. 6. Die regressielyne, wat as liniêr beskou kan word, is met behulp van die volgende formule akkuraat bepaal:

$$Y = r \cdot \frac{S_y}{S_x} \cdot x - r \cdot \frac{S_y}{S_x} \cdot \bar{x} + \bar{y} \quad (23, p.107),$$

(waar x = I.K.,
 y = eksamenpunt,
 r = waargenome korrelasie tussen x en y ,
 S_x, S_y = standaardafwykings van x en y ,
 \bar{x}, \bar{y} = gemiddelde waardes van x en y)

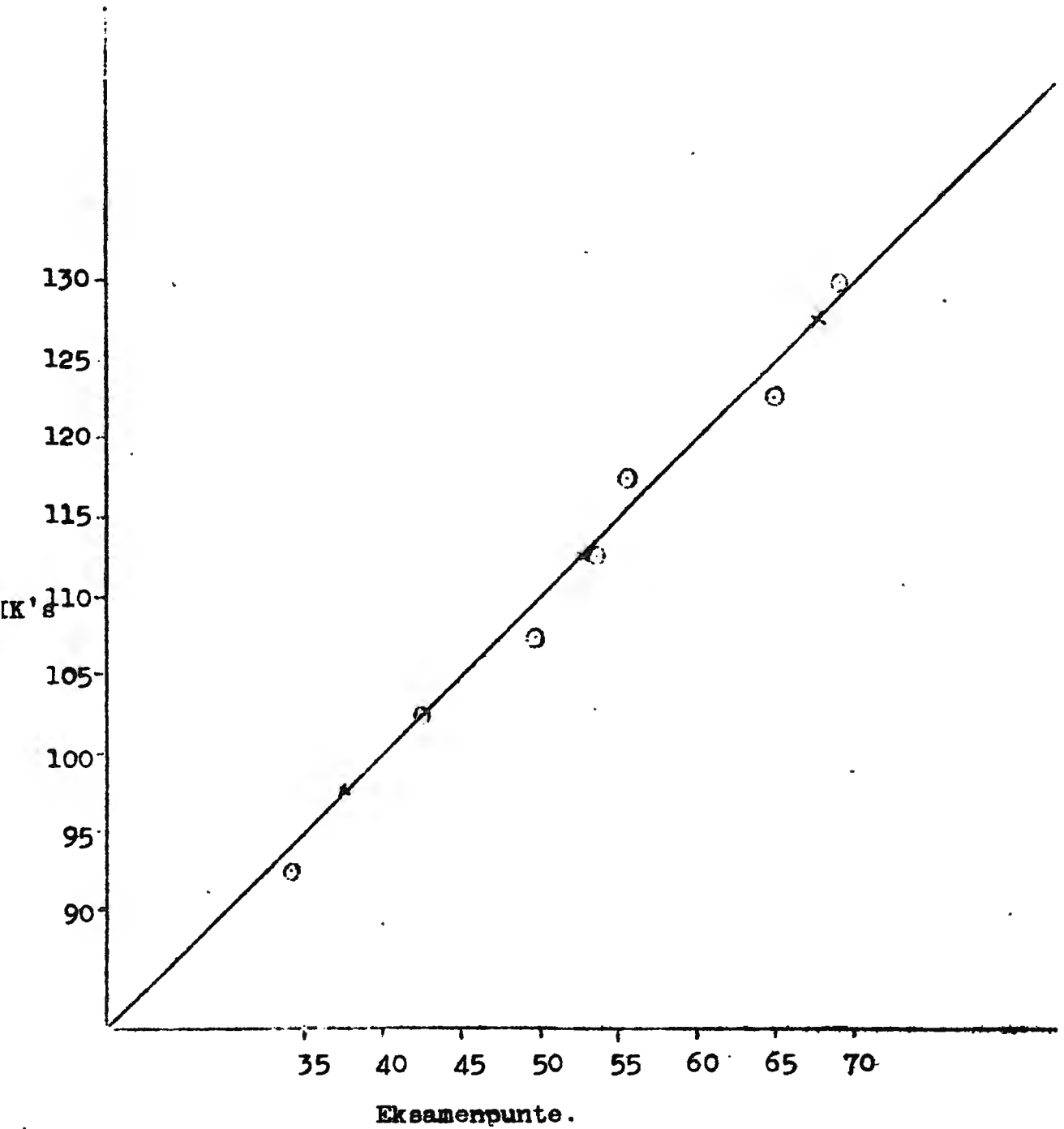
en word in Figure 3 en 4 (pp. 47&48) aangegee. Van

die/.....

Regressie van Eksamenpunte op I K's

Std. 6, Groep A

$$r = .78$$

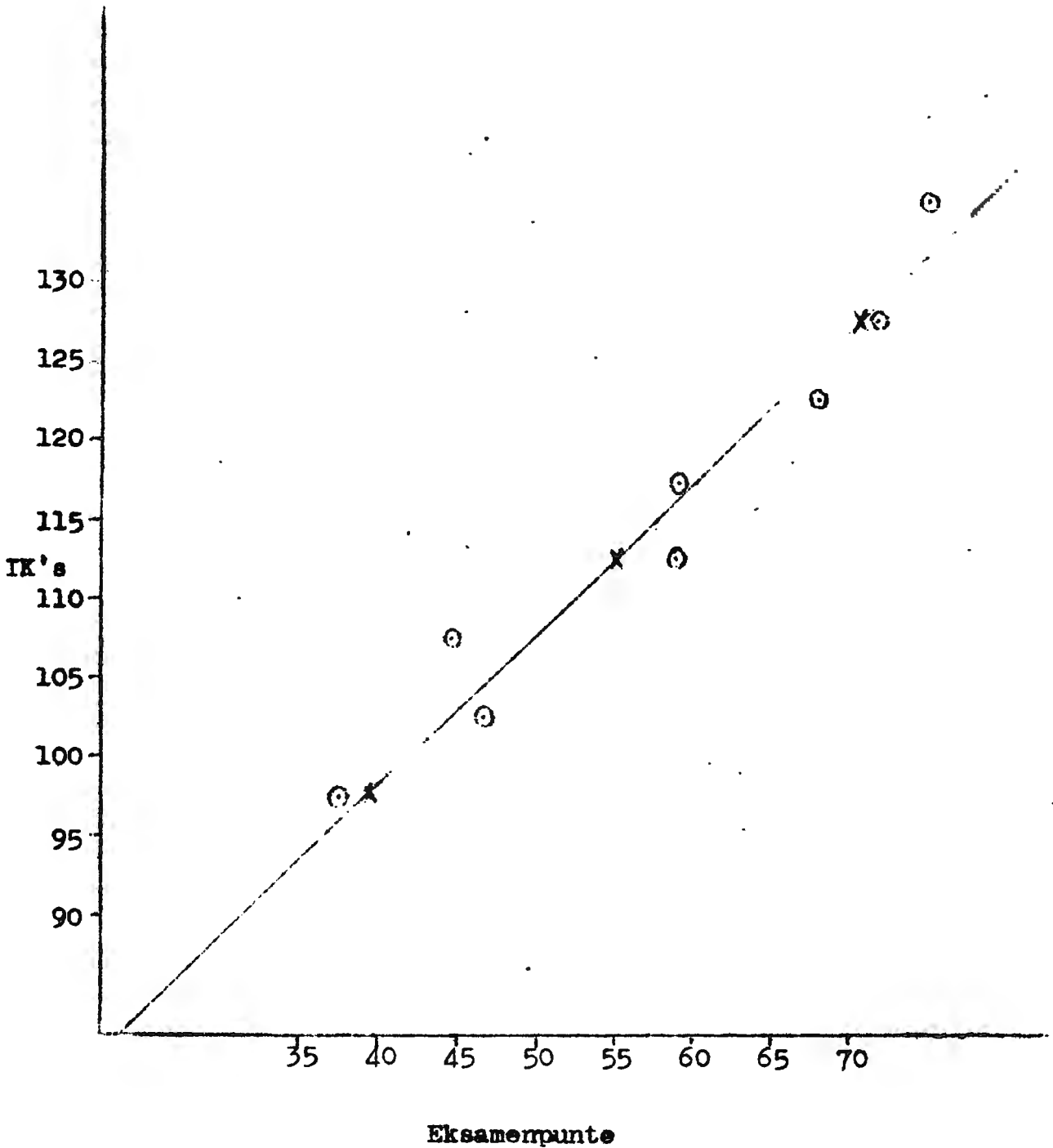


Figuur 3.

Regressie van Examenpunte op I K's

Std. 6, Groep B

$$r = .77$$



Figuur 4

die regressielyn kon nou vir elke I.K. die „verwagte“ eksamenpunt afgelees word. Afwykings van die verwagte eksamenpunte is voorgestel deur die verskil; behaalde eksamenpunt minus verwagte eksamenpunt met behoorlike algebraïese teken. 'n Positiewe afwyking dui dus 'n prestasie bō die verwagte, en 'n negatiewe afwyking 'n prestasie onder die verwagte vlak aan. Hierdie afwyking is vervolgens met die vraelystellings gekorreleer. Die koëffisiënte word in Tabel 12 aangegee.

Korrelasiekoëffisiënte tussen afwykings van verwagte eksamenprestasies en vraelyspunte Std. 6.

	N	r
Std.6, groep A	85	-.063
Std.6, groep B	85	-.233

TABEL 12

Die negatiewe korrelasie is soos verwag is ('n hoë vraelystelling gepaard met 'n groot negatiewe afwyking, en omgekeerd), maar slegs die B-groep se koëffisiënt was beduidend by die .05-waarskynlikheidsvlak.

Hierdie manier van geldigheidsbepaling is tot die Std.6-groepe bepaal omdat die korrelasie tussen eksamenpunte en I.K.'s in Stds. 7 en 8 waarskynlik nie hoog genoeg is om verwagte eksamenpunte met behulp van die regressie van eksamenpunte op I.K.'s te bepaal nie (Sien Tabelle 8 en 9, pp.43-44).

Volgens albei metodes vir die bepaling van die geldigheid van die Persoonlike Vraelys skyn dit dus of die vraelys min geldigheid vir die meet van wanaangepastheid besit. Dit is natuurlik moontlik dat die kriterium vir wanaangepastheid wat gebruik is nie

toepaslik/.....

toepaslik is nie, en dat die vraelys hoër geldigheid mag vertoon as dit met 'n geldige kriterium vergelyk word. Opsommend kan dus gesê word dat die Persoonlike Vraelys 'n instrument met hoë betroubaarheid maar lae of hoogstens onbekende geldigheid vir die vasstelling en meet van emosionele wanaangepastheid is.

2. Die AG-toets.

Die AG-toets vir Verstandelike Wakkerheid, wat in besonderhede deur Morkel (24) bespreek word, is gedurende die Tweede Wereldoorlog deur die Aanlegtoetsafdeling van die Suid-Afrikaanse Lugmag ontwikkel. Die opsteller het as model gebruik die Otis-selfadministrende (hoër-)greepsverstandtoets, soos dit deur die Suid-Afrikaanse Buro vir Opvoedkundige Navorsing gewysig is. In vergelyking met lg. toets het die AG-toets meer items van 'n wiskundige en relasionele aard, en is sekere dubbelsinnige en onduidelike items weggelaat. Die AG-toets is met sukses vir die keuring van nie-gematrikuleerde lugbemanning en grondpersoneel gebruik. Die norms vir die AG-toets (asook vir die A/8- en die M-toets) wat in hierdie ondersoek gebruik is, is gebaseer op die toetsprestasies van 1637 rekrute, wat gedurende die tydperk Februarie 1947 tot April 1948 getoets is, en wat 'n gemiddelde skoolstanderd van 6.8 gehad het.¹ Die omvorming van rou punte tot standaardpunte neem nie ouderdomsverskille tussen proefpersone in aanmerking nie - 'n feit wat by die ontleding van AG- gegewens vir verskillende skoolstanderds in ag geneem moet word.

Daar is twee parallelle Afrikaanse vorms van die toets, nl. AG1 en AG2. Elk bestaan uit 60 items wat vir die twee vorms so na as moontlik gelykwaardig

¹ Ongepubliseerde materiaal in navorsingslêer Ki/51/7 van die N.I.P.N. ten/.....

ten opsigte van aard en moeilikheidswaarde is. Daar is verskillende soorte items, o.a. sinonieme, klassifikasie, sillogismes, syferreeks, analogieë, ens., wat naastenby in stygende moeilikheidsvolgorde gerangskik is.

Elke vraag word beantwoord deur 'n letter of 'n syfer tussen die hakies aan die einde van die vraag te plaas. Op die eerste bladsy van die toets word uitvoerige instruksies en 6 voorbeelde aangegee. Drie van die voorbeelde is reeds beantwoord om die manier van beantwoording aan te toon, die orige drie moet deur die proefpersoon self beantwoord word. Voor die aanvang van die eintlike toets moet die antwoorde op hierdie drie vrae deur die proefnemer gekontroleer word om seker te maak dat die proefpersoon die instruksies begryp het.

Die proefpersone word aangesê om vinnig en akkuraat te werk, en nie onnodig tyd te verspil met items wat hulle te moeilik vind nie. Die toetsduur waarvoor die norms vasgestel is, en wat in hierdie ondersoek gebruik is, is 30 minute.

Die feit dat die AG-toets se inhoud van 'n betreklike bekende, skoolse aard is, en dat dit vir 'n nie-gematrikuleerde populasie en op die model van 'n „selfadministrerende“ toets ontwikkel is, maak dit taamlik waarskynlik dat dit 'n maklike toets, wat inhoud sowel as instruksies betref, vir Std. 6-8 leerlinge sal wees. Die toets het waarskynlik ook 'n taamlike mate van gesigsgeldigheid en belangwekkendheid.

Om/.....

Om bogenoemde redes, en omdat ervaring in die N.I.P.N. daarop gedui het dat dit as 'n goeie aanvangstoets beskou kan word, is besluit om toets AG in die eerste eksperimentele battery in te sluit.

3. Die A/8-toets.

Die A/8-toets is 'n toets van rekenkundige vermoë en prestasie wat gedurende die Tweede Wêreldoorlog deur die Aanlegstoetsafdeling van die Suid-Afrikaanse Lugmag ontwikkel is. Dit bestaan uit 62 items, waaronder eenvoudige voorbeelde van die vier hoofbewerkings, nl. optel, aftrek, vermenigvuldig en deel, hoofbewerkings met breuke, omvorming van gewone breuke tot desimale breuke en omgekeerd, verhoudingsprobleme, vormleerprobleme, ens. Hierdie toets het ook geblyk geldig te wees in die militêre keuringsituasie. Die vorms wat in hierdie ondersoek gebruik is, is op die toetsprestasies van dieselfde groep rekrute as wat vir die AG- en M- toetse gebruik is, gebaseer (p. 50). Die A/8-toets het 'n tydsgrens van 40 minute.

Daar bestaan twee parallelle Afrikaanse vorms van die toets, nl. A/8/1 en A/8/2. Die instruksies in die toets is so volledig dat die toets feitlik selfadministrerend is.¹ Om hierdie rede, asook om die skoolse aard van die toets en die feit dat prestasie op rekenkundige toetse skynbaar nie so nadelig deur wanaangepastheid beïnvloed word nie (p. 13), is besluit om die A/8-toets in die tweede eksperimentele battery in te sluit.

4. Die /.....

¹ Vergelyk die duur van die instruksies vir die verskillende toetse met mekaar (Tabelle 44 en 45, p. 88).

4. Die M-toets¹

Die M-toets, soos in hierdie ondersoek gebruik, is die minder-bekende „Progressive Matrices 1943” van Raven, wat dit spesiaal vir die keuring van offisiere in die Britse leër gedurende die Tweede Wêreldoorlog ontwerp het. Dit bestaan uit 'n reeks van 38 visueel-aangebode probleme in 'n boekie met een item per bladsy. Op die boonste helfte van die bladsy is 'n groot reghoekige ontwerp wat 'n matrys (matrice) genoem word. Uit die onderste regterhandse kwadrant van elke matrys is 'n stuk verwyder. Op die onderste helfte van die bladsy word 8 matrysfragmente, almal van geskikte formaat om in die oop ruimte op die matrys in te pas, maar met verskillende ontwerpe daarop, aangegee. Elk van hierdie 8 fragmente is 'n moontlike keuse om die matrys te voltooi, en die proefpersoon moet dié fragment kies waarvan die ontwerp die korrekte is om die matryspatroon te voltooi.

'n Oefenstel van 4 items gaan die toetsreeks vooraf, en die beantwoording van die oefenstel word deur die proefnemer gekontroleer voordat die toets begin, om te verseker dat die proefpersoon die instruksies verstaan.

Daar is geen deurlopende beginsel van probleemoplossing nie - die beginsel van oplossing moet vir elke probleem weer opnuut afgelei word. Terwyl hierdie toets oorspronklik as 'n „kragtoets” (d.w.s. sonder tydsgrens) bedoel en gebruik is, is dit in die N.I.P.N. vir 'n rekrutepopulasie² gestandaardiseer met 'n tydsgrens van 20 minute, wat ook nog tot die moeilikheid

van/.....

¹ Met M-toets word in hierdie verhandeling deurgaans „Progressive Matrices 1943” bedoel.

² Dieselfde as in die geval van die AG- en A/8-toets, p. 50

van die toets bygedra het. Sover vasgestel kon word is daar nog geen verslag oor die „Progressive Matrices 1943” gepubliseer nie, hoewel 'n hele aantal publikasies oor die makliker vorm, nl. „Progressive Matrices 1938” verskyn het. Aangesien die M-toets vir 'n superieure populariteitsbedoel is, kan aangeneem word dat dit, veral as spoedtoets, veel moeiliker is as „Progressive Matrices 1938”, en dat sulke kenmerke van die 1938-variant, soos gevoeligheid vir emosionele wanaangepastheid, spanning, vermoedens, ens., nog meer beklemtoon sal wees in die 1943-vorm. Omdat dit 'n toets is wat visuele oordeelsvermoë vereis, die instruksies moeilik is, en dit in die M.I.P.N. algemeen as 'n besonder ongeskikte aanvangstoets beskou is, is die M-toets in albei eksperimentele batterye ingesluit.

D. DIE PROEFPERSONE.

1. Algemeen.

Omdat bekendheid met die toetse in die eksperimentele batterye 'n onbekende invloed op die toetsprestaties, en die houdings teenoor die toetssituasie sou hê, kon dieselfde groep proefpersone nie vir die vergelyking van die invloed van verskillende aanbiedingsvolgordes gebruik word nie. Om dieselfde rede kon die kies van eksperimentele groepe wat gelykwaardig ten opsigte van prestasies op die toetse in die eksperimentele batterye sou wees, nie op die basis van 'n voorafgaande toepassing van die betrokke toetse geskied nie. 'n Prosedure vir gelykwaardigmaking van die eksperimentele groepe, onafhanklik van die prestasies op die toetse in die eksperimentele batterye, was dus nodig. Hiertoe moes van alle toepaslike gegewens oor die proefpersone gebruik gemaak word.

Faktore wat die toetsprestasies moontlik kan beïnvloed is onder andere ouderdom, hoeveelheid skoolopleiding ontvang, sosiale, ekonomiese en beroepsagtergrond, huistaal en geslag. As die eksperimentele groepe ten opsigte van al sulke hoedanighede gelykwaardig gemaak kon word, sou hulle waarskynlik ook gelykwaardig wees ten opsigte van toetse in die eksperimentele batterye. Dit is egter uiters moeilik om groepe gelykwaardig ten opsigte van meer as een veranderlike te kies, en daarom is dit nodig dat slegs die belangrikste bepalende faktore vir toetsprestasies as beginsels vir gelykwaardigmaking gebruik word. Dit is een van die redes waarom van skoolkinders as proefpersone gebruik gemaak is - vir hulle was daar 'n aanduiding van die algemene intelligensie in terme van die intelligensiekoëffisiënt (I.K.), beskikbaar, wat hoogs waarskynlik ook 'n goeie aanduiding van die prestasievermoë in die toetse in die eksperimentele toetsbatterye sou gee. Daar is dus aangeneem dat die I.K.'s hoog met die prestasies in die betrokke toetse sou korreleer - hierdie vermoede is later bevestig (PP-44-45, 66-67).

Hoewel dit nie heeltemal korrek is om bv. soos Wechsler (39, p. 109) aan te neem dat groepe wat gelykwaardig is ten opsigte van prestasies op een toets, ook gelykwaardig moet wees ten opsigte van prestasies op enige ander toets wat hoog met die eerste korreleer, is 'n soortgelyke metode tog in hierdie ondersoek gebruik. Die verskil was egter dat die groepe proefpersone wat gelykwaardig gekies is uit dieselfde skool geneem is; dat hulle dus by benadering homogeen sal wees ten opsigte van huistaal, ouderdom, sosiale agtergrond en ervaring in toetssituasies. Verder het die

kies/.....

kies van gelykwaardige groepe op grond van twee belangrike veranderlikes, nl. skoolstanderd¹ en I.K.'s plaasgevind, en was al die proefpersone van dieselfde geslag. Die gevolgtrekking dat so 'n prosedure gelykwaardige groepe proefpersone sal oplewer is baie meer aanneemlik as wanneer proefpersone net ten opsigte van 'n enkele veranderlike gelykwaardig gekies word. 'n Onderzoek wat na afloop van die hoofonderzoek uitgevoer is het dan ook aangetoon dat die gekose groepe wel gelykwaardig was (pp. 68-70).

2. Eerste eksperiment.

Vir die eerste eksperiment is die Std. 6 - 8 seuns van 'n Afrikaanse hoërskool in Johannesburg, hierna „1^e skool” genoem, gebruik. Nadat die seuns in 'n aanpassingsklas uitgeskakel is, was daar nog 487 seuns beskikbaar, waarvan die I.K.'s van 456 bekend was. Besonderhede oor die samestelling van hierdie groep proefpersone word in Tabel 13 aangegee.

Samestelling van oorspronklike groep proefpersone in eerste eksperiment.

	Std. 6	Std. 7	Std. 8	Totaal Std. 6 - Std. 8
Aantal in Standaard.	197	164	126	487
Aantal waarvan I.K.'s bekend is.	183	152	121	456

TABEL 13

Die 456 seuns waarvan die I.K.'s bekend was is vervolgens binne elke skoolstanderd in twee groepe, A en B, wat so na as moontlik identies ten opsigte van I.K.-verspreiding was, verdeel. Die twee groepe kon nog nie volkome identies gemaak word nie omdat

daar/.....

¹ In 'n vroeëre ondersoek het die skrywer vasgestel dat daar 'n sterk verband tussen toetsprestasie en hoogste skoolstanderd geslaag bestaan (13, pp. 5, 9-10).

daar in die I.K.- verspreiding binne elk van die drie skoolstanderds verskeie klasintervalle was met 'n onewe aantal proefpersone daarin. Die „oortolliges” is ook nie verwyder nie, omdat hulle moontlik in die plek van persone wat op die dag van die eksperiment afwesig sou wees, gebruik kon word. Sodoende het die totale A-groep uit 244 seuns en die totale B-groep uit 243 seuns bestaan voordat die eksperimentele ondersoek uitgevoer is. Ten tyde van die uitvoering van die eerste eksperiment was 25 seuns afwesig, en is een uitgeskakel omdat hy 'n immigrant was en Afrikaans nog nie behoorlik magtig was nie. Die verspreiding van I.K.'s vir die A- en B-groepe binne elke skoolstanderd van al die proefpersone wat die toetse afgele het, (en waarvan die I.K.'s bekend was), is vervolgens bepaal, en die minimum aantal proefpersone is volgens alfabetiese volgorde van name verwyder sodat die oorblywende A- en B- groepe binne elke skoolstanderd 'n identiese I.K.-verspreiding vertoon het. Die I.K.-verspreidings, met die ooreenkomstige gemiddeldes en standaardafwykings, vir elke skoolstanderd, en ook van die totale gelykwaardige groepe proefpersone wat uiteindelik in die eerste eksperiment gebruik is, word in Tabel 14 aangegee.

In die eerste eksperiment was daar dus uiteindelik 85 Std. 6- seuns, 66 Std. 7- seuns, en 53 Std. 8- seuns, tesame 204, in Groep A, en 'n ooreenkomstige aantal, met presies dieselfde skoolstanderdsamestelling en I.K.- verspreiding binne elke skoolstanderd in Groep B.

I.K. - verspreidings, -gemiddeldes, en - Standaard-afwykings van die finale gelykwaardige A- en B-groepe, eerste eksperiment.

Klasintervalle	F R E K W E N S I E S.			Totaal Std.6- Std.8.
	Std.6	Std.7	Std.8	
135 - 139	1	1		2
130 - 134	2	2		4
125 - 129	4	2	1	7
120 - 124	8	3	6	17
115 - 119	11	14	12	37
110 - 114	12	15	10	37
105 - 109	17	17	18	52
100 - 104	12	10	4	26
95 - 99	7	2	2	11
90 - 94	7			7
85 - 89	3			3
80 - 84	1			1
N	85	66	53	204
Gemiddeldes	108.529	111.697	111.528	110.333
Standaard-afwykings.	11.454	8.159	6.748	9.506

TABEL 14

3. Tweede eksperiment.

Vir die tweede eksperiment is die Std. 6 - 8 seuns van 'n ander Afrikaanse hoërskool in Johannesburg, hierna „2^e skool” genoem, gebruik. Net soos in die eerste eksperiment is die proefpersone vocraf in twee groepe, A en B wat naastenby gelykwaardig ten opsigte van skoolstandersamestelling en I.K.- verspreiding binne elke skoolstander was, verdeel, en na afloop van die eksperiment is die minimum aantal proefpersone verwyder om die A- en B- groepe presies gelykwaardig ten opsigte van skoolstandersamestelling en I.K.- verspreiding binne elke skoolstander te laat.

Oorspronklik was daar 290 seuns vir die ondersoek beskikbaar, en van 284 van hulle was die I.K.'s bekend. Die samestelling van hierdie groep word in Tabel 15 aangetoon. Van hierdie groep was daar ten tyde van die uitvoering van die eksperiment 25 afwesig.

Samestelling van oorspronlike groep proefpersone in tweede eksperiment

	Std.6	Std.7	Std.8	Totaal Std.6- Std.8
Aantal in Standerd	149	81	60	290
Aantal waar- van I.K.'s bekend is.	147	79	58	284

TABEL 15

Die I.K.-verspreidings, met die ooreenkomstige gemiddeldes en standaardafwykings, vir elke skoolstander, en ook van die totale gelykwaardige groepe proefpersone wat uiteindelik in die tweede eksperiment gebruik is, word in Tabel 16 aangegee. Daar was dus 64 Std. 6 - seuns, 31 Std. 7- seuns en 23 Std. 8- seuns, tesame 118,

in Groep A, en 'n ooreenkomstige aantal, met presies dieselfde skoolstandaardsamestelling en I.K.- verspreiding binne elke skoolstanderd in Groep B. Die eksperimentele gegewens is van hierdie gelykwaardige groepe proefpersone verkry.

I.K.-verspreiding, -gemiddeldes en-standaard-afwykings van die finale gelykwaardige A- en B-groepe, tweede eksperiment.

Klasintervalle	Std.6	Std.7	Std.8	Totaal Std.6-Std.8
120 - 124	1		1	2
115 - 119	4	3	1	8
110 - 114	11	2	3	16
105 - 109	9	8	9	26
100 - 104	16	10	7	33
95 - 99	12	3	2	17
90 - 94	6	3		9
85 - 89	3	2		5
80 - 84	2			2
N	64	31	23	118
Gemiddeldes	102.469	103.226	106.739	103.356
Standaard-afwykings.	8.826	7.606	5.651	8.173

TABEL. 16

E. DIE PROEFNEMERS.

Sewe vroulike proefnemers van die N.I.P.N. wat opleiding in toetstoedieningsmetodes ontvang het, en praktiese ervaring in die toediening van toetse AG, A/B

en/.....

en M gehad het, is in die ondersoek gebruik. Nog 7 helpsters, een by elk van die proefnemers, was met die uitdeel en opneem van toetse, en met die toesighou behulpsaam. Terwille van konstante toetsomstandighede is deurgaans van dames gebruik gemaak - die skrywer het algemene toesig gehou oor die toetsprosedure, maar self geen toetse toegedien nie.

F. DIE UITVOERING VAN DIE TOETSPROGRAM.

1. Algemeen.

Die voorwaardes waaronder toestemming verkry is om in die betrokke skole te toets, sowel as praktiese oorwegings soos versekering van konstante toetsomstandighede, voorkoming van die uitlek van inligting en die moontlike houdingsveranderinge wat dit teweeg sou bring, ens., het bepaal dat al die proefpersone in 'n bepaalde eksperiment almal gelyktydig getoets moes word. Klaskamcrakkommodasie sowel as die vereistes vir doeltreffende toediening van groeptoetse het verder bepaal dat elke proefnemer se groep nie groter as ongeveer 35 persone moes wees nie. Die twee gelykwaardige groepe van elke eksperiment moes dus in subgroepe onderverdeel word, en vir elke subgroep moes 'n proefnemer in 'n lokaal afgesonderd word. Naamlyste van die subgroepe is opgestel en vooraf aan die skoolhoof gegee, sodat so min tyd as moontlik verlore sou gaan met die oorskakeling van die skoolroetine na die eksperimentele situasie. Die doel en metode van ondersoek is vooraf breedvoerig met die twee skoolhoofde en met elke proefnemer wat in die ondersoek gebruik is, bespreek, sodat hulle presies sou weet waarom die ondersoek op die bepaalde wyse uitgevoer is. Elke proefnemer het ook 'n eksemplaar van die naamlys van haar subgroep gekry,

sy moes voor die aanvang van die toetsessie seker maak dat sy die regte proefpersone daar het, en sy moes die aanwesiges en afwesiges op die lys aandui. Op elke naamlys is die volgorde waarin die toetse aan die bepaalde subgroep gegee moes word, aangedui. Elke proefnemer moes ook, in samewerking met haar helpster, aandui hoe lank die instruksies vir elk van die twee toetse geduur het, en wat die totale toepassingsduur van die Persoonlike Vraelys was. Vir hierdie doel is aan elke proefnemer 'n stophorlosie gegee.

Op versoek van die skrywer het elke skoolhoof kort voor die aanvang van die toetsessie die proefpersone meegedeel dat hulle intelligensietoetse sou aflê. Op hierdie wyse is getrag om 'n sterk motivering vir die doen van die toetse by die proefpersone te bewerkstellig.

2. Eerste eksperiment.

Die eerste eksperiment is op Vrydag 18 Mei 1951 uitgevoer. Die gelykwaardige A- en B-groepe was elk in 7 subgroepe, A1 tot A7, en B1 tot B7 ingedeel. Daar was dus 14 subgroepe en net 7 proefnemers, sodat elke proefnemer twee subgroepe sou moes hanteer. Hierdie beperking het voordelig geblyk te wees omdat dit beheer van die differensiële invloed van verskillende proefnemers op die toetsprestasies gebied het. Die subgroepe is nl. só geplaas dat die A1- en B1- subgroepe se lokale direk langs mekaar was, die A2- en B2- subgroepe langs mekaar, ens. Elke proefnemer het sodoende 'n A- en 'n B- subgroep van naastenby gelykwaardige samestelling gekry om te hanteer, en enige besondere invloed wat sy op die toetsprestasies kon uitoefen bv. deur onduidelike spraak, sou albei gelykwaardige groepe tot dieselfde mate tref.

Die/.....

Die A-subgroepe het die toetse M en AG1 in daardie volgorde afgelê, en die B- subgroepe dieselfde toetse in die volgorde AG1 - M. Albei groepe het na voltooiing van die twee toetse onmiddellik die Persoonlike Vraelys beantwoord. Elke proefnemer het eers die instruksies vir die M- toets aan die A- subgroep gegee, daarna, terwyl die helpster en die klasonderwyser by die A- subgroep toesig gehou het, het sy die instruksies vir die AG-toets aan die ooreenstemmende B- subgroep gaan gee. Aangesien die M-toets 'n tydsgrens van 20 minute, en die AG-toets 'n tydsgrens van 30 minute het, kon sy albei groepe ook laat ophou na verloop van die vasgestelde tyd. Voordat die B-subgroepe se tyd vir die AG- toets verstreke was, het die A-subgroepe reeds die M-toets voltooi, en ook die instruksies vir die AG toets ontvang. Sorg is gedra dat proefpersone nie alleen gelaat is nie, en dat die proefnemers nie onder druk instruksies gegee het nie. Die voorgaande prosedure is gevolg om die totale duur van die eksperimentele battery so kort moontlik te maak, sodat daar nie ernstige inbreuk op die skoolwerkzaamhede gemaak sou word nie.

3. Tweede eksperiment.

Die tweede eksperiment is op Maandag 21 Mei 1951 uitgevoer. Die gelykwaardige A- en B- groepe is elk in 4 subgroepe, A1 tot A4, en B1 tot B4 ingedeel, sodat daar dus 8 subgroepe was. Twee van die subgroepe, nl. A4 en B4, was heelwat kleiner as die res, en hulle is in rotasie deur dieselfde proefnemer getoets, soortgelyk met die reëling in die eerste eksperiment. Die originele 6 subgroepe is elk deur 'n afsonderlike proefnemer getoets. Die A-subgroepe het die toetse M en A/8/1 in daardie volgorde, en die B- subgroepe het dieselfde toetse in die omgekeerde volgorde

afgelê/....

afgelê. Onmiddellik na voltooiing van die twee toetse het al die proefpersone die Persoonlike Vraelys beantwoord.

4. Hertoetsing met die parallele toetse.

Op Vrydag 10 Augustus 1951 is al die beskikbare proefpersone van die eerste eksperiment met die AG2-toets hertoets, en op dieselfde dag is ook al die beskikbare proefpersone van die tweede eksperiment met die A/3/2-toets hertoets. Aangesien al die proefpersone in 'n skool dus dieselfde toetse moes aflê, is hulle nie weer in die oorspronklike subgroepe ingedeel nie, maar is die gewone klasindeling behou. Van die proefpersone wat tydens die hertoetsing afwesig was, is verskeie by later geleentheid getoets. Aangesien die oorspronklike gelykwaardige groepe nie voltallig teenwoordig was by die hertoetsing met die parallele toetse nie, moes opnuut A- en B- groepe wat gelykwaardig was ten opsigte van skoolstanderdsamestelling en I.K.- verspreiding binne elke skoolstanderd gekies word, sodat die geldigheid van die gelykwaardigmakingsprosedure teen die prestasies op die parallele toetse getoets kon word.

G. STATISTIESE TEGNIEKE WAT VIR DIE VERWERKING VAN GEGEWENS GEBRUIK IS.

1. Korrelasie. In alle gevalle waar die koëffisiënt van korrelasie tussen twee veranderlikes bepaal moes word, is van die Pearson-produkmoment-formule gebruik gemaak. Volgens subjektiewe beoordeling kon die kantverspreiding in al die gevalle as naastenby normaal, en die regressielyste as reguit beskou word (23, p. 120).

Meervoudige/...

Meervoudige korrelasiekoëffisiënte wat slegs drie veranderlikes behels het, (twee veranderlikes se meervoudige korrelasie met 'n derde), is met behulp van 'n formule wat deur Guilford gegee word (15, p.385) bereken. Meervoudige korrelasiekoëffisiënte wat meer as drie veranderlikes behels het, is volgens Aitken se spilverdigtingsmetode (pivotal condensation) bereken (22, pp. 32-37).

2. Statistiese toets vir beduidendheid. Verskille tussen twee gemiddeldes of twee waargenome persentasies, onafhanklik of gekorreleer, is deur middel van die verhouding tussen waargenome verskil en die standaardfout van die betrokke statistiek vir beduidendheid getoets (23, pp.59-82). Die kritiese verhoudings wat met bepaalde waarskynlikheidswaardes saamhang, is van Fisher en Yates se tabelle (10, p.32) afgelees. Om die beduidendheid van waargenome korrelasiekoëffisiënte te bepaal, is die kritiese waardes van die korrelasiekoëffisiënte wat met bepaalde steekproefgroottes en bepaalde waarskynlikheidswaardes saamhang, van Fisher en Yates se tabelle (10, p.46) afgelees. Om die statistiese beduidendheid van die verskille tussen die gemiddeldes van individue wat in 'n tweevoudige klassifikasie gerangskik is mee te toets, is van die tegniek van tweevoudige variansieontleding soos deur Kendall (20, pp.220-223) beskrywe, gebruik gemaak. Kritiese waardes van die variansieverhouding is van McNemar se Tabel F (23, pp.353-355) afgelees.

H. DIE/....

H. DIE EKSPERIMENTELE RESULTATE.

1. Gelykwaardigheid van die A- en B-groepe ten opsigte van toetsprestasies.

In hierdie afdeling sal aangetoon word dat die prosedure wat vir die gelykwaardigmaking van die A- en B-groepe in die twee eksperimente gebruik is, doeltreffend was. Soos in die bespreking van die geldigheid van die Persoonlike Vraelys geblyk het, was daar 'n korrelasie van die orde .5 tot .8 tussen I.K.'s en A/8-punte, en van die orde .3 tot .5 tussen I.K.'s en M-punte in die eerste eksperiment (sien Tabelle 7 - 9, pp. 43-44). Die interkorrelasies tussen I.K.'s, A/8-, M- en Vraelyspunte soos in die tweede eksperiment gevind, word in Tabelle 17 - 19 aangegee.

Korrelasiematriks, Std. 6, tweede eksperiment.

	I.K.'s	A/8	M	Vraelys
I.K.'s	/	.665	.429	-.234
A/8	.574	/	.544	-.243
M	.554	.536	/	-.011
Vraelys	-.079	-.298	-.102	/

A-Groep. N = 64

B-Groep. N = 64

Tabel 17.

Die skynbaar-afwykende waardes wat vir die interkorrelasie tussen I.K.'s en A/8 vir Std. 8, Groep A, en tussen I.K.'s en M vir Std. 8, Groep B, verkry word, kan waarskynlik verklaar word op grond van die groter steekproefvariasie waaraan die gegewens vir die kleiner groepe in die tweede eksperiment onderhewig is.

Die/....

Die korrelasie tussen I.K.'s en A/8 kan gevolglik as van die orde .5 tot .7 beskou word, en dié tussen

Korrelasiematriks, Std. 7, tweede eksperiment.

	I.K.'s	A/8	M	Vraelys
I.K.'s		.603	.312	-.038
A/8	.561		.183	.012
M	.355	.199		.025
Vraelys	-.440	.048	-.277	

A-Groep N = 31

B-Groep N = 31.

Tabel 18.

Korrelasiematriks, Std. 8, tweede eksperiment.

	I.K.'s	A/8	M	Vraelys
I.K.'s		.559	.087	-.178
A/8	-.056		.263	-.195
M	.310	.293		-.063
Vraelys	.034	-.012	.061	

A-Groep N = 23

B-Groep N = 23

Tabel 19.

I.K.'s en M as van die orde .3 tot .5 (soos in die eerste eksperiment). Dit blyk uit Tabelle 14 en 16, pp. 58 en 60, dat daar 'n inperking van die variasiebreedte van I.K.'s in die hoër standerds is. As hier-

die/....

hierdie inperking van talent in ag geneem word by die berekening van die interkorrelasies, word aansienlike hoër koëffisiënte as die voorgaande verkry, soos die skrywer elders aangetoon het (14, pp. 48-49). Hoewel die veronderstelde verband tussen I.K.'s en toetsprestasies dus bevredigend aangetoon is, waarborg dit nog nie dat die A- en B-groepe ook ten opsigte van toetsprestasies gelykwaardig sal wees nie. Klaarblyklik kon prestasies in toetse in die eksperimentele batterye ook nie gebruik word om enersyds te bepaal of toetsvolgordes 'n beduidende invloed op die toetsprestasies uitoefen, en andersyds aan te toon dat die gekose A- en B-groepe ook ten opsigte van toetsprestasies gelykwaardig is nie. Vir hierdie doel kon egter die prestasies in die parallelle toetse AG2 en A/8/2 ontleed word. Nadat die A- en B-groepe van die twee eksperimente deur weglating van 'n minimum aantal proefpersone weer ten opsigte van skoolstandersamestelling en I.K.-verspreiding binne elke skoolstanderd volkome gelykwaardig gemaak is, is die gemiddelde punte op die parallelle toetse vir hierdie gelykwaardige groepe bepaal. Die resultate word in Tabela 20 en 21 aangegee.

AG2-punte, opvolgingstoets, eerste eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6- Std. 8
A-Groep	\bar{x}	52.983	59.581	61.750	57.504
	S	8.740	5.968	4.900	7.965
	N	75	61	50	186
B-Groep	\bar{x}	51.783	60.401	62.250	57.423
	S	3.219	5.894	5.564	8.304
	N	75	61	50	186

Tabel 20.

A/8/2-punte, opvolgingstoets, tweede eksperiment.

		Std. 6.	Std. 7.	Std. 8.	Totaal Std. 6- Std. 8.
A-Groep	\bar{x}	49.593	50.129	57.450	51.307
	S	9.105	6.194	7.237	8.626
	N	56	28	21	105
B-Groep	\bar{x}	48.878	54.236	56.021	51.736
	S	9.196	9.563	9.016	9.769
	N	56	28	21	105

Tabel 21.

Variansieontleding van hierdie gegewens het aangetoon dat die verskille tussen die gemiddelde toetsprestasies in nie een van die twee gevalle beduidend is nie. Aangesien geen parallele vorm van die M-toets beskikbaar was nie, kon die gelykwaardigheid van die A- en B-groepe ten opsigte van prestasies op die M-toets nie op 'n soortgelyke wyse aangetoon word nie. Die ouderdomme van al die proefpersone in die twee eksperimente is korrek tot die naaste tiende van 'n jaar - op die dag van die eksperiment - bereken, en aangesien hierdie veranderlike nie eksplisiet by die keuse van gelykwaardige groepe in ag geneem is nie (maar waarskynlik hoog met die skoolstanderd korreleer), is dit as betekenisvol beskou dat 'n variansieontleding van die gegewens, soos in Tabelle 22 en 23 aangegee, getoon het dat ook die verskille tussen die gemiddelde ouderdomme van die A- en B-groepe in nie een van die twee eksperimente statisties beduidend was nie.

Gemiddelde/....

Gemiddelde ouderdomme van proefpersone in eerste eksperiment op 18 Mei 1951.

		Std. 6.	Std. 7.	Std. 8.	Totaal Std. 6-Std. 8.
A-Groep	\bar{x}	14.098	14.669	15.462	14.637
	S	.871	.708	.872	.982
	N	85	66	53	204
B-Groep	\bar{x}	13.863	14.669	15.670	14.593
	S	.896	.623	.757	1.067
	N	85	66	53	204

Tabel 22.

Gemiddelde ouderdomme van proefpersone in tweede eksperiment op 21 Mei 1951.

		Std. 6.	Std. 7.	Std. 8.	Totaal Std. 6-Std. 8.
A-Groep	\bar{x}	14.114	15.261	15.874	14.747
	S	.793	.701	.433	1.017
	N	64	31	23	118
B-Groep	\bar{x}	14.370	15.422	15.680	14.902
	S	.745	.828	.444	.936
	N	64	31	23	118

Tabel 23.

Die direkte en omstandighedsgetuienis skyn dus die gevolgtrekking dat die A- en B-groepe gelykwaardig was ten opsigte van prestasievermoë in die AG- en M-, en A/8- en M-toetse, te regverdig.

2. Die subjektiewe beoordeling van die toetsituasie.

By die beplanning van die eksperimentele ondersoek is aangeneem dat die aard van die taak wat in die AG- en A/8-toetse gestel is, sowel as die instruksies/....

instruksies vir hierdie toetse, vir die proefpersone betreklik maklik sou wees, terwyl die M-toets, vanweë sy oorsprong, en die aard van die instruksies en taak, vir hulle beslis moeilik sou wees. In hierdie afdeling sal vasgestel word tot watter mate hierdie aannames geregtvaardig is.

Die vrae in die vraelys wat direk op die subjektiewe beoordeling van die toetsituasie betrekking het, is die volgende:

- Vraag b. Watter toets was vir jou die moeilikste?
- Vraag c. Van watter toets het jy die meeste gehou?
- Vraag d. In watter toets dink jy het jy die beste gedoen.

Hierdie drie vrae staan op die voorblad van die vraelys, en moes deur die invul van M of AG/A/8 beantwoord word. Die volgende vier vrae, in die „Ja-Nee-?“-vorm verskyn op die tweede bladsy van die vraelys, en het ook op die toetsituasie betrekking.

- Vraag 1. Voel dit vir jou of jy goed gedoen het in toets AG (of A/8)?
- Vraag 2. Voel dit vir jou of jy goed gedoen het in toets M?
- Vraag 3. Het jy moeg gevoel nadat jy die eerste toets voltooi het?
- Vraag 4. Hinder dit jou baie as jy dink dat jy dalk sleg gedoen het in hierdie toets?

(a) Die moeilikheid van die toets.

Hoewel dit waagsaam is om die antwoorde op die vrae in 'n vraelys letterlik te vertolk, is die antwoorde op vrae b en d vir die vergelyking van die subjektiewe moeilikheidswaardes van die toets gebruik.

Dit is veronderstel dat 'n bewus-deurdenkte antwoord op die vraag: „Watter toets was vir jou die moeilikste?“, ook die antwoord op die vraag: „In watter toets dink jy het jy die beste gedoen?“ sou bepaal, en wel só, dat „M“ as antwoord op die eerste vraag tot „AG“ (of „A/8“) as antwoord op die tweede vraag sou lui, en omgekeerd.

Die aantal kere wat elke alternatief op elke vraag gekies is, tesame met die persentasie van die totaal wat dit voorstel, word vir die totale A- en B-groepe in elk van die twee eksperimente in Tabelle 24 en 25 aangegee.

Antwoorde op vrae b en d, eerste eksperiment.

Vraag	Groep	M		AG		Uitgelaat	
		N	%	N	%	N	%
b	A N=204	112	54.9	91	44.6	1	0.5
	B N=204	170	83.3	34	16.7		
d	A N=204	69	33.8	135	66.2		
	B N=204	49	24.0	155	76.0		

Tabel 24.

Antwoorde/....

Antwoorde op vrae b en d, tweede eksperiment.

Vraag	Groep	M		A/8		Albei	
		N	%	N	%	N	%
b	A N=118	31	26.3	87	73.7		
	B N=118	84	71.2	34	28.8		
d	A N=118	71	60.2	46	39.0	1	0.8
	B N=118	50	42.4	68	57.6		

Tabel 25.

Die eenvoudigste hipotese, volgens algemene statistiese gebruik, wat aangaande die relatiewe moeilikheid van die toetse opgestel kan word, is dat hulle ewe moeilik gevind word deur die proefpersone. Daarvolgens kan verwag word dat die persentasie proefpersone wat sê M is die moeilikste toets nie beduidend van 50% sal verskil nie - so dus ook vir die persentasie wat AG (of A/8) as die moeilikste toets aandui. In Tabel 24, vir vraag b, verskil die persentasie wat die M-toets as die moeilikste beskou het in die geval van die A-groep nie beduidend van 50% nie, maar in die geval van die B-groep is die persentasie beduidend groter as 50% ($p < .001$)². Die persentasie persone wat op vraag d geantwoord het dat hulle dink hulle het die beste in toets AG gedoen -

dus/.....

² D.w.s. die waarskynlikheid dat 'n verskil soos die waargenome deur kansvariasie sal ontstaan, is kleiner as 1 in 1000.

dus die swakste in toets M- is vir beide Groepe A en B beduidend groter as 50% ($p < .001$). Dit is opmerklik dat die verskil tussen die A- en B-groepe se antwoorde op vraag b beduidend is ($p < .001$), en so ook vir vraag d ($p < .05$). In albei gevalle dui 'n groter persentasie van die B-groep (wat M tweede gedoen het) M as die moeilikste toets aan. Op grond van hierdie getuienis kan aangeneem word dat albei groepe proefpersone in die eerste eksperiment die M-toets moeiliker as die AG-toets gevind het, en dat dit veral met die B-groep, wat die M-toets tweede gedoen het, die geval was.

In Tabel 25, vir vraag b, is die persentasie wat die M-toets as die moeilikste beskou het in die geval van die A-groep beduidend kleiner as 50% ($p < .001$), en in die geval van die B-groep beduidend groter as 50% ($p < .001$). Die persentasie persone wat op vraag d geantwoord het dat hulle dink hulle het die beste in toets A/8 gedoen - dus die swakste in toets M - is vir Groep A beduidend kleiner as 50% ($p < .05$), terwyl dit vir Groep B nie beduidend groter as 50% is nie. Daar is dus groot verskille tussen die A- en B-groepe se antwoorde. Vir vraag b is die verskil in persentasie van die A- en B-groepe wat 'n bepaalde antwoord kies beduidend ($p < .001$), en so ook in die geval van vraag d ($p < .01$).

In die tweede eksperiment vind ons dus dat die A-groep die A/8-toets as die moeilikste beskou, terwyl die B-groep die M-toets as die moeilikste beskou.

Ten/.....

Ten opsigte van die ooreenkoms tussen antwoorde op vrae b en d vind ons in albei eksperimente dat as die persentasie van 'n groep wat M as antwoord kies op vraag b groter as 50% is, dan is die persentasie van dieselfde groep wat M as antwoord kies op vraag d kleiner as 50%, soos verwag is. Tog is vir beide die A- en B-groepe, in beide eksperimente, die verskil tussen die persentasie van 'n groep wat 'n bepaalde antwoord op vraag b kies, en die persentasie van dieselfde groep wat die ooreenstemmende, verwagte antwoord op vraag d kies, beduidend ($p < .001$). Klaarblyklik het vraag b en vraag d virminstens sommige van die proefpersone nie dieselfde betekenis gehad nie.

(b) Die gewildste toets.

Die aantal en persentasie persone wat elke alternatief as antwoord gekies het op die vraag: „Van watter toets het jy die meeste gehou?” word vir die twee eksperimente afsonderlik in Tabelle 26 en 27 aangegee.

Antwoorde op vraag c, eerste eksperiment.

Vraag	Groep	M		AG	
		N	%	N	%
a	A	97	47.5	107	52.5
	B	73	35.8	129	63.2

Tabel 26.

~~Antwoorde/....~~

Antwoorde op vraag c. Tweede eksperiment.

Vraag	Groep	M		A/8	
		N	%	N	%
c	A	76	64.4	42	35.6
	B	56	47.5	62	52.5

Tabel 27

Die eenvoudigste statistiese hipotese is dat toets M en AG (of A/8) ewe gewild is, en dat elke alternatief dus deur 50% van proefpersone gekies sal word.

In die geval van die eerste eksperiment (Tabel 26) is die persentasie proefpersone wat meer van toets AG as van toets M gehou het vir beide Groep A en Groep B groter as 50%, maar dis slegs in geval van Groep B statisties beduidend groter as 50% ($p < .001$). Die verskil in persentasies van die A- en B-groepe wat die meeste van 'n bepaalde toets (hetse AG of M) gehou het, is statisties beduidend ($p < .02$).

In die geval van die tweede eksperiment is die persentasie van die A-groep wat die meeste van toets M gehou het beduidend groter as 50% ($p < .01$), terwyl die persentasie van die B-groep wat die meeste van toets M gehou het, nie beduidend kleiner as 50% was nie. Die verskil tussen die persentasies van die A- en B-groepe wat die meeste van 'n bepaalde toets (hetse A/8 of M) gehou het is beduidend ($p < .01$).

Op grond van die voorgaande resultate kan aangeneem word dat die M-toets in die eerste eksperiment minder gewild was as die AG-toets, terwyl die M-toets in die tweede eksperiment skynbaar meer gewild as die

A/8-.....

A/8-toets was. Dit is opmerklik dat die B-groepe (wat toets M tweede gedoen het) in albei eksperimente minder van die F-toets gehou het as die A-groepe.

(c) Eie prestasie in die toetse.

Uit die oorsig oor die literatuur het geblyk dat die subjektiewe beoordeling van eie toetsprestasie as 'n belangrike faktor in die toetssituasie beskou is (p. 10). Voordat die toetsprestasies wat met verskillende volgordes verkry is, vergelyk word, is dit dus nodig om vas te stel hoe die verskillende groepe hulle eie prestasies beoordeel het. Die antwoorde op vrae 1 en 2 in die vraelys is met die oog hierop ontleed¹. Die eksperimentele gegewens verskyn in Tabelle 28 en 29.

Antwoorde op vrae 1 en 2, eerste eksperiment.

Vraag	Groep	Ja		Nee		?		Uitgelaat	
		N	%	N	%	N	%	N	%
1	A N=204	155	76.0	37	18.1	12	5.9		
	B N=204	165	80.9	27	13.2	12	5.9		
2	A N=204	106	51.9	75	36.8	23	11.3		
	B N=204	95	46.6	68	33.3	38	18.6	3	1.5

Tabel 28.

Antwoorde/....

¹ Vraag 1 lui: Voel dit vir jou of jy goed gedoen het in toets AG (of A/8)?

Vraag 2 lui: Voel dit vir jou of jy goed gedoen het in toets M?

Antwoorde op vrae 1 en 2, tweede eksperiment.

Vraag	Groep	Ja		Nee		?		Uitgelaat	
		N	%	N	%	N	%	N	%
1	A N=118	52	44.1	57	48.3	9	7.6		
	B N=118	77	65.2	33	28.0	8	6.8		
2	A N=118	78	66.1	25	21.2	15	12.7		
	B N=118	60	50.9	47	39.8	11	9.3		

Tabel 29.

Wat die gegewens vir die eerste eksperiment (Tabel 28) betref, het dit geblyk dat vir sowel groep A as groep B 'n beduidende kleiner persentasie van die proefpersone gedink het dat hulle goed gedoen het in toets M as wat gedink het dat hulle goed gedoen het in toets AG ($p < .001$ in albei gevalle). Die verskille tussen die A- en B-groepe se antwoorde op vrae 1 en 2 is, hoewel nie statisties-beduidend nie, in dieselfde rigting as die voorafgaande resultate oor die subjektiewe beoordeling van die toetssituasie. Daar is nl. meer van die B- as van die A-groep-proefpersone wat gedink het dat hulle goed gedoen het in toets AG, en meer proefpersone van die A- as van die B-groep het weer gedink dat hulle goed gedoen het in toets M.

Uit die gegewens vir die tweede eksperiment (Tabel 29) het geblyk dat 'n beduidende groter persentasie van die A-groep gedink het dat hulle goed gedoen het in toets M as wat gedink het dat hulle goed gedoen het in toets A/8 ($p < .001$). In die geval van die B-groep het/.....

het 'n beduidende groter persentasie van die proefpersone gedink hulle het goed gedoen in toets A/8 as wat gedink het dat hulle goed gedoen het in toets M ($p < .001$). Wat betref die verskille in reaksies van die A- en B-groepe dink 'n beduidende groter persentasie van die B- as van die A-groep dat hulle goed gedoen het in toets A/8 ($p < .01$), terwyl 'n beduidende groter persentasie van die A- as van die B-groep dink dat hulle goed gedoen het in die M-toets ($p < .02$).

Dit is opmerklik dat die antwoorde op vrae 1 en 2 tot dieselfde bevinding lei as die antwoorde op vrae b en d.

(d) Vermoeidheid.

Die antwoorde op vraag 3: „Het jy moeg geword nadat jy die eerste toets voltooi het?” word in Tabelle 30 en 31 aangetoon.

Antwoorde op vraag 3, eerste eksperiment.

Vraag	Groep	Ja		Nee		?		Uitgelaat	
		N	%	N	%	N	%	N	%
3	A N=204	7	3.4	194	95.1	2	1.0	1	0.5
	B N=204	11	5.4	189	92.6	2	1.0	2	1.0

Tabel 30.

Antwoorde op vraag 3, tweede eksperiment.

Vraag	Groep	Ja		Nee		?		Uitgelaat	
		N	%	N	%	N	%	N	%
3	A N=118	10	8.5	105	89.0	2	1.7	1	0.8
	B N=118	8	6.8	109	92.4				

Tabel 31.

In/....

In nie een van die twee eksperimente verskil die persentasie Ja-antwoorde van die A-groep beduidend van dié van die B-groep nie. Dit is opmerklik dat slegs 'n klein persentasie proefpersone in beide eksperimente geantwoord het dat hulle moeg gevoel het na voltooiing van die eerste toets.

(e) Motivering. Om 'n aanduiding te kry van hoe sterk die motivering was om goed te doen in die toetse, of watter persoonlike betekenis die proefpersone aan die toetsituasie geheg het, is die antwoorde op vraag 4: „Hinder dit jou baie as jy dink dat jy dalk sleg gedoen het in hierdie toetse?” ontleed. Die gegewens is in Tabelle 32 en 33 saamgevat.

Antwoorde op vraag 4, eerste eksperiment.

Vraag	Groep	Ja		Nee		?		Uitgelaat	
		N	%	N	%	N	%	N	%
4	A N=204	131	64.2	61	29.9	10	4.9	2	1.0
	B N=204	132	64.7	59	28.9	12	5.9	1	0.5

Tabel 32.

Antwoorde op vraag 4, tweede eksperiment.

Vraag	Groep	Ja		Nee		?		Uitgelaat	
		N	%	N	%	N	%	N	%
4	A N=118	46	39.0	66	55.9	6	5.1		
	B N=118	58	49.2	53	44.9	5	4.2	2	1.7

Tabel 33.

In nie een van die twee eksperimente was daar 'n beduidende verskil tussen die persentasies van die A- en die B-groepe wat "Ja" op vraag 4 geantwoord het nie. Daar is egter 'n opmerklieke verskil, beduidend by die .01-waarskynlikheidsvlak, tussen die persentasies proefpersone in die eerste en tweede eksperimente wat hierdie vraag bevestigend beantwoord het. As 'n letterlike vertolking aan hierdie resultaat gegee kan word, sal dit waarskynlik lui dat dit die proefpersone in die eerste eksperiment veel meer erns was met die toetsituasie, dat daar 'n groter mate van ego-betrekking in hulle geval was. Daarom daar so 'n verskil in waardering van die toetsituasie sou ontstaan is nie duidelik nie.

3. Die Invloed van Toetsvolgorde.

Daar is reeds aangetoon dat die A- en B-groepe in elk van die twee eksperimente waarskynlik as gelykwaardig ten opsigte van potensiele prestasies in toetse in die eksperimentele battery beskou kan word. Verder is ook aangetoon wat die subjektiewe beoordelings van verskillende aspekte in die eksperimentele toetsituasie was, en hoe die Persoonlike Vraelys se geldigheid vir die meet van emosionele wanaangepastheid nie bevredigend vasgestel kon word nie. In die volgende paragrawe sal die toetsprestasies wat met verskillende volgordes verkry is, vergelyk word, en sal die verband tussen 'n hoë vraelyspunt en moontlike gevoeligheid vir toetsvolgorde ondersoek word.

(a) Toetsprestasies in die eerste eksperiment.

Die gemiddelde punte wat deur die proefpersone van die eerste eksperiment met verskillende toetsvolgordes op toetse AG en M behaal is, word in Tabelle 34 en 35 aangegee.

Gemiddelde/.....

Gemiddelde AG-punte, eerste eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6- Std. 8
A-groep	\bar{x}	48.332	55.783	57.450	53.112
	S	7.120	6.533	5.408	7.697
	n	85	66	53	204
B-groep	\bar{x}	46.685	56.314	58.016	52.744
	S	8.026	6.560	6.635	8.879
	n	85	66	53	204

Tabel 34.

Gemiddelde M-punte, eerste eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6- Std. 8
A-groep	\bar{x}	50.215	55.632	56.601	53.626
	S	9.564	8.240	6.569	8.930
	n	85	66	53	204
B-groep	\bar{x}	51.274	56.465	60.092	55.244
	S	7.920	8.129	7.985	8.789
	n	85	66	53	204

Tabel 35

Die verskille tussen die A- en B-groepe se gemiddeldes op sowel toets AG as toets M is nie statisties beduidend nie. Hierdie resultaat is op twee maniere verkry. Ten eerste is met behulp van die t-toets vasgestel dat die verskille tussen die AG-gemiddeldes van die totale A- en B-groepe, en tussen die M-gemiddeldes van die totale A- en B-groepe, nie statisties beduidend is nie. Daarna is gegewens vir albei toetse met behulp van die tegniek van tweevoudige variansieontleding behandel, en geen beduidende verskil tussen die gemiddelde toetsprestasies van die A- en B-groep is gevind nie. Tog is die verskille tussen die M-gemiddeldes in al drie skoolstandards in die verwagte rigting - die B-groep, wat die M-toets/....

M-toets ná die AG-toets afgelê het, het 'n hoër M-gemiddelde as die A-groep, wat die M-toets eerste afgelê het. Daarteenoor het die A-groep slegs in Std. 6 'n hoër AG-gemiddelde as die B-groep.

(b) Toetsprestasies in die tweede eksperiment.

Die gemiddelde A/8- en M-punte van die A- en B-groepe in die tweede eksperiment word in Tabelle 36 en 37 aangegee.

Gemiddelde A/8-punte, tweede eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6- Std. 8.
A-groep	\bar{Y}	50.419	48.579	57.233	51.264
	S	8.278	6.440	6.990	8.174
	n	64	31	23	118
B-groep	\bar{Y}	49.481	53.579	55.494	51.730
	S	8.602	8.586	10.605	9.372
	n	64	31	23	118

Tabel 36

Die verskille tussen die gemiddelde prestasies van die A- en B-groepe in die A/8- sowel as die M-toets is nie statisties-beduidend nie, beide volgens die t-toets en die tegniek van tweevoudige variansieontleding.

Gemiddelde M-punte, tweede eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6- Std. 8.
A-groep	\bar{Y}	44.716	49.547	47.233	46.475
	S	11.690	10.300	10.883	11.373
	n	64	31	23	118
B-groep	\bar{x}	44.559	50.031	53.320	47.704
	S	8.606	10.616	8.679	9.869
	n	64	31	23	118

Tabel 37.

Hoewel/....

Hoewel onbeduidend is die verskil in M-gemiddeldes van die totale A- en B-groepe in die verwagte rigting - die groep wat die M-toets na die A/8-toets afgelê het, d.w.s. groep B, presteer effens beter as die ander groep. Die B-groep presteer ook effens beter as die A-groep op die A/8-toets, hoewel ook hierdie verskil onbeduidend is.

(c) Proefpersone met 'n hoër vraelyspunt.

Nieteenstaande die feit dat die geldigheid van die Persoonlike Vraelys nie oortuigend aangetoon kon word nie, is besluit om die invloed van toetsvolgorde op toetsprestasie te bepaal vir die groep wat 'n vraelystelling van 20 of hoër verkry het. Die keuse van hierdie punte, wat ongeveer een kwart tot 'n helfte van 'n standaardafwyking onder die gemiddelde vraelyspunt van die verskillende groepe was, was grootliks willekeurig. Die keuse van 'n hoër kritiese vraelyspunt sou die grootte van die steekproef egter aansienlik verminder.

In Tabelle 38 en 39 word die gemiddelde AG- en M-punte van al die proefpersone in die eerste eksperiment wat 'n vraelyspunt van 20 of hoër behaal het, aangegee. Die AG-gemiddeldes van die A- en B-groepe is baie na aan mekaar, met die B-groep s'n effens, maar nie beduidend nie, hoër. Die gemiddelde M-punte van die B-groep

B-groep/...

Gemiddelde AG-punte van proefpersone met 'n vraelyspunt van 20 of hoër, eerste eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6- Std. 8.
A-groep	\bar{X}	46.918	55.018	57.450	52.363
	S	6.937	5.650	5.080	7.632
	n	47	37	31	115
B-groep	\bar{X}	46.177	56.736	58.479	52.957
	S	8.214	6.144	6.943	9.115
	n	55	49	34	138

Gemiddelde M-punte van proefpersone met 'n vraelyspunt van 20 of hoër, eerste eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6- Std. 8.
A-groep	\bar{x}	49.857	55.288	53.160	55.320
	S	9.218	8.603	5.875	6.704
	n	47	37	31	115
B-groep	\bar{x}	51.359	56.238	60.979	55.675
	S	7.609	7.398	8.092	8.572
	n	55	49	34	138

Tabel 39

B-groep is egter in al drie standerds hoër as dié van die A-groep. Tweevoudige variansientleding het aangetoon dat die verskille beduidend is ($P < .03$).

In Tabelle 40 en 41 word die gemiddelde A/8- en M-punte van al die proefpersone met 'n vraelyspunt van 20 of hoër in die tweede eksperiment aangetoon. Die totale B-groep

Gemiddelde A/8-punte van proefpersone met 'n vraelyspunt van 20 of hoër, tweede eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6 - Std. 8
A-groep	\bar{x}	49.962	49.572	57.300	50.286
	S				8.264
	n	45	18	13	74
B-groep	\bar{x}	48.588	54.510	53.012	50.605
	S				8.924
	n	48	19	17	84

Tabel 40

Gemiddelde M-punte van proefpersone met 'n vraelyspunt van 20 of hoër, tweede eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6 - Std. 8
A-groep	\bar{x}	44.194	46.978	47.015	45.288
	S				11.481
	n	43	18	13	74
B-groep	\bar{x}	44.429	49.653	51.500	47.093
	S				9.548
	n	48	19	17	84

Tabel 41

B-groep se gemiddelde prestasies in die A/8- en M-toets is in albei gevalle effens hoër as die ooreenstemmende gemiddelde toetsprestasies van die totale A-groep, maar nie een van die twee verskille is statisties-beduidend nie. As die aantalle in die tweede eksperiment egter groter was, is dit byna seker dat die verskil tussen die A- en B-groepe se gemiddelde M-punte beduidend sou wees. Daar is twee redes vir hierdie oortuiging. Ten eerste is die absolute verskille tussen die M-gemiddeldes van die A- en die B-groepe in die twee eksperimente byna dieselfde, en ten tweede het die B-groep in elke standard 'n hoër M-gemiddelde as die A-groep behaal.

Omdat net dié proefpersone wat 'n vraelyspunt van 20 of hoër behaal het vir hierdie ontleding gebruik is, is die aantalle in die ooreenstemmende A- en B-groepe nie gelyk soos in al die ander ontledings.

(d) Die invloed van toetsvolgorde op die gemiddelde vraelyspunte.

Dit is opmerklik dat in albei eksperimente 'n groter aantal proefpersone uit die B- as uit die A-groep 'n vraelyspunt van 20 of hoër behaal het. Toe hierdie aspek ondersoek is, is dan ook gevind dat die gemiddelde vraelyspunt van die B-groep in albei eksperimente hoër was as dié van die A-groep, hoewel die verskille nie statisties-beduidend was nie. Die betrokke gegewens word vir die twee eksperimente afsonderlik in Tabelle 42 en 43 aangegee.

Gemiddelde vraelyspunte, eerste eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6 - Std. 8
A-groep	\bar{x}	22.347	23.560	22.179	22.696
	S	9.703	11.211	10.183	10.354
	n	85	66	53	204
B-groep	\bar{x}	25.076	24.045	22.632	24.108
	S	10.716	9.254	10.231	10.184
	n	35	66	53	204

Tabel 42

Gemiddelde vraelyspunte, tweede eksperiment.

		Std. 6	Std. 7	Std. 8	Totaal Std. 6 - Std. 8
A-groep	\bar{x}	25.562	22.905	22.826	24.314
	S	10.352	8.030	10.536	9.953
	n	64	31	23	118
B-groep	\bar{x}	26.688	23.839	25.174	25.669
	S	11.307	11.118	8.860	10.904
	n	64	31	23	118

Tabel 43

Dit is opmerklik dat, hoewel die verskille nie beduidend is nie, die B-groep in elke standaard en in albei eksperimente 'n hoër gemiddelde vraelyspunt as die A-groep het.

(e) Toetsvolgorde en instruksie-duur.

Terwyl dit in die eerste eksperiment duidelik geword het dat toetsvolgorde die toetsprestasies van sekere proefpersone (nl. dié met 'n vraelyspunt van 20 of hoër) beduidend beïnvloed, en die verskille in toetsprestasies in die tweede eksperiment ook in die verwagte rigting is, moet een verklaringswyse, wat vroeër reeds genoem is, nl. die duur van die instruksies met verskillende toetsvolgordes nog ondersoek word. Die hoër M-prestasies van die B-groepe mag daaraan te wyte wees dat dié proefpersone, omdat hulle na aflegging van die eerste toets (AG of A/8) minder vreemd en skaam gevoel het, meer geredelik vrae gestel het terwyl die instruksies vir toets M gegee is. Die instruksieduur tot die naaste minuut vir elke toets in elke subgroep word vir die twee eksperimente afsonderlik in Tabelle 44 en 45 aangegee.

Instruksieduur/...

Instruksieduur (in minute), eerste eksperiment.

A-GROEP	TOETS AG	TOETS M	VRAELYS (TOTAAL)	B-GROEP	TOETS AG	TOETS M	VRAELYS (TOTAAL)
A1	5	8	22	B1	5	8	20
A2	7	12	20	B2	5	11	22
A3	5	15	23	B3	7	14	23
A4	6	13	19	B4	7	10	20
A5	4	11	19	B5	5	7	16
A6	4	6	19	B6	5	6	24
A7	4	5	20	B7	3	5	23
Gemid. duur, Totale A-groep	5.00	10.00	20.29	Gemid. duur, Totale B-groep	5.29	8.71	21.14

Tabel 44

In die eerste eksperiment (Tabel 44) is die instruksietye streng vergelykbaar, omdat A- en B-subgroepe met ooreenstemmende nommers (bv. A3 en B3) deur dieselfde proefnemer getoets is. In die tweede eksperiment (Tabel 45) is slegs die A4- en B4-subgroepe deur dieselfde toetsers getoets, en is die gemiddelde instruksietye dus nie streng vergelykbaar nie, omdat praattempo kan wissel van persoon tot persoon. Nogtans kom die gemiddelde instruksietye

Instruksieduur (in minute), tweede eksperiment.

A-GROEP	TOETS A/8	TOETS M	VRAELYS (TOTAAL)	B-GROEP	TOETS L/8	TOETS M	VRAELYS (TOTAAL)
A1	1	11	20	B1	2	15	21
A2	2	10	20	B2	2	10	20
A3	1	8	18	B3	1	7	17
A4	1	14	13	B4	1	11	15
Gemid. duur Totale A-Groep	1.25	10.75	17.75	Gemid. duur Totale B-Groep	1.50	10.75	18.25

Tabel 45

vir toets M in die twee eksperimente en vir die A- en B-groepe, asook die instruksietye vir toets AG en A/8 vir die betrokke A- en B-groepe goed ooreen. Die verskille wat daar wel is, dui daarop dat 'n toets as aanvangstoets waarskynlik 'n langer toedieningsduur het as wanneer dit later in die battery voorkom. Die verbetering in gemiddelde M-punte wat vir die B-groepe waargeneem is, kan dus nie aan 'n langer instruksieduur (omstandiger verduideliking of meer vrae) toegeskryf word nie. Weens die klein aantalle waarnemings is geen toets vir beduidendheid van verskille in toetsduur uitgevoer nie. Aangesien die verskille in die teenoorgestelde rigting is as wat op grond van verklarings in terme van instruksieduur verwag is, is beduidendheidstoets ook nie nodig nie.

Die totale toedieningstye, d.w.s. die tydverloop vandat die instruksies begin totdat die laaste proefpersoon in die subgroep klaar is, word vir die Persoonlike Vraelys aangegee. Dit is opmerklik dat die proefpersone in die tweede eksperiment, asook die proefpersone in die A-groepe, onderskeidelik korter gemiddelde vraelystye as dié van die tweede eksperiment en die B-groepe het, hoewel die verskille moontlik nie beduidend is nie.

I. SAMEVATTING VAN DIE EKSPERIMENTELE RESULTATE.

Ter wille van oorsigtelikheid, en om die samehang van die verskillende bevindings te bepaal, word die resultate van die eksperimentele ondersoek kortliks hier saamgevat. Die numeriese en alfabetiese aanduiding van die verskillende hoofde stem met dié van die vorige afdeling (H. DIE EKSPERIMENTELE RESULTATE) ooreen, sodat verwysing na die volledige bespreking gerieflik kan plaasvind.

1. Gelykwaardigheid van die A- en B-groepe ten opsigte van toetsprestasies.

Die prestasies van die A- en B-groepe van die eerste eksperiment in toets AG2, en van die A- en B-groepe van die tweede eksperiment in toets A/8/2; verder, die vergelyking van die gemiddelde ouerdomme van die A- en B-groepe in albei eksperimente;

uiteindelik/...

uiteindelik, die prestasies van die totale A- en B-groepe van albei eksperimente in die toetse van die eksperimentele batterye, dui daarop dat die prosedure wat vir die verkryging van gelykwaardige groepe gebruik is, in albei eksperimente doelmatig was.

2. Die subjektiewe beoordeling van die toetsituasie.

(a) Die moeilikheid van die toetse. In die eerste eksperiment het die proefpersone die M-toets moeiliker as die AG-toets gevind. Hierdie resultaat het heelwat sterker op die voorgrond getree in die geval van die B-groep (wat die M-toets ná die AG-toets afgelê het) as in die geval van die A-groep. In die tweede eksperiment het die A-groep, (wat die A/8 toets ná die M-toets afgelê het), die A/8-toets as die moeilikste beskou, terwyl die B-groep (wat die M-toets na die A/8-toets afgelê het), toets M as die moeilikste beskou het.

(b) Die gewildste toets. Die M-toets was in die eerste eksperiment minder gewild as die AG-toets, terwyl die M-toets in die tweede eksperiment skynbaar die gewildste was. Dit is opmerklik dat die B-groepe (wat in albei eksperimente die M-toets tweede gedoen het) minder van die M-toets gehou het as die A-groepe.

(c) Eie prestasie in die toetse. In die eerste eksperiment het minder van die proefpersone gedink dat hulle goed gedoen het in toets M as wat gedink het dat hulle goed gedoen het in toets AG. Dit was veral met die B-groep die geval. In die tweede eksperiment het meer van die proefpersone in die A-groep gedink dat hulle goed gedoen het in toets M as wat gedink het hulle het goed gedoen in toets A/8, terwyl meer van die proefpersone van die B-groep gedink het dat hulle goed gedoen het in toets A/8 as wat gedink het dat hulle goed gedoen het in toets M. In albei eksperimente word dus gevind dat minder proefpersone van die B-groepe as van die A-groepe gedink het dat hulle goed gedoen het in toets M.

(d) Vermoedheid. In albei eksperimente het slegs 'n paar proefpersone aangedui dat hulle moeg gevoel het na voltooiing van die eerste toets. Daar was geen beduidende verskille tussen die A- en B-groepe nie.

(e) Motivering. Die eerste eksperiment se proefpersone was skynbaar sterker gemotiveer om goed te doen in die toets as die tweede eksperiment s'n. In albei eksperimente was die verskille tussen die resultate van die A- en B-groepe onbeduidend.

3. Die Invloed van Toetsvolgorde.

(a) Toetsprestasies in die eerste eksperiment. Daar was geen beduidende verskille tussen die totale A- en B-groepe se gemiddelde prestasies nie, sowel wat die AG- as wat die M-toets betref. Tog is die B-groep se gemiddelde M-prestasie in elke standaard hoër as die A-groep s'n. Daarteenoor het die A-groep slegs in een standaard, nl. std. 6, 'n hoër AG-gemiddelde as die B-groep.

(b) Toetsprestasies in die tweede eksperiment. Die verskille tussen die totale A- en B-groepe se gemiddelde prestasies was nie beduidend nie vir sowel die A/8- as die M-toets. Ook hier is die B-groep se M-gemiddelde effens hoër as die A-groep s'n. Ook die B-groep se A/8-gemiddelde is effens hoër as die A-groep s'n.

(c) Proefpersone met 'n hoër vraelyspunt. In die eerste eksperiment is gevind dat as net dié proefpersone met 'n vraelyspunt van 20 of hoër geneem word, dan is die M-gemiddelde van die B-groep beduidend hoër as dié van die A-groep. In die tweede eksperiment was die B-groep se M-gemiddelde vir al drie skoolstanderds hoër as dié van die A-groep. Hoewel die verskil tussen die M-gemiddeldes van die A- en die B-groepe nie beduidend was nie, is aangevoer dat met groter aantalle proefpersone dit waarskynlik wel beduidend sou wees. In albei eksperimente het die B-groepe ook effens hoër gemiddeldes as die A-groepe op die ander toetse (respektiewelik AG en A/8) behaal.

(d) Toetsvolgorde en gemiddelde vraelyspunt. 'n Ontleding van die vraelyspunte het aangetoon dat die B-groep in albei eksperimente in elk van die standerds 'n hoër gemiddelde vraelyspunt as die A-groep behaal het. Hoewel die verskil statisties onbeduidend is, is dit 'n betekenisvolle feit dat die resultaat vir al 6 pare gelykwaardige subgroepe geld.

(e) Toetsvolgorde en instruksieduur. In nie een van die twee eksperimente was daar opmerklike verskille in die instruksieduur as dieselfde toets eers as 'n aanvangstoets en dan as 'n tweede toets toegepas is nie. Die waargenome verskille dui, indien ongetwyfel, aan dat die instruksies van 'n toets as aanvangstoets langer duur as wanneer dieselfde toets later in die reeks toegedien word.

Watter gevolgtrekkings op grond van hierdie bevindings gemaak kan word, en tot watter mate dit die vooropgestelde hipoteses en bestaande menings bevestig, wysig of repudieer sal in die volgende hoofstuk aangetoon word.

Hoofstuk IV

'N KRITIESE-SAMEVATTENDE BESKOUING VAN DIE ONDERSOEK

A. ALGEMEEN

1. Inleidings. 'n Kritiese bespreking van die resultate van hierdie studie moet in die eerste plek aantoon tot watter mate die doelstellinge van die ondersoek bereik is. Dit moet ook aantoon hoe die bevindinge met die gangbare menings en vooropgesette hipoteses vergelyk, en hoe die beskouings gewysig moet word. In feitlik elke eksperimentele ondersoek word daar ook waarnemings gemaak wat nie spesifiek op die probleem wat ondersoek word, betrekking het nie. In sover sulke bevindinge van algemene belang vir die sielkunde as vakwetenskap mag wees, moet hulle ook gerapporteer word. Laastens moet by die verskillende fases van die bespreking aange-
toon word in watter rigting verdere navorsing nodig en waarskynlik vrugbaar sal wees.

Om die kritiese bespreking te oriënteer word die doelstelling van hierdie studie weereens kortliks gestel. In die eksperimentele ondersoek moes

- (a) Vasgestel word of, en tot watter mate, die twee verskillende aanbiedingsvolgordes van twee potlood- en papiertoeetse die toetsprestasies van proefpersone beïnvloed;
- (b) As vasgestel kon word dat toetsvolgorde toetsprestasies beduidend beïnvloed, bepaal word wat die bydras van sekere emosionele faktore tot die waargenome verskille in toetsprestasies is. (Sien pp. 4-5, 18)

2. Kritiese samevatting van die gangbare beskouings oor toetsvolgorde en die invloed van emosionele faktore op toetsprestasies. (pp. 7-17)

Uit die bestudeerde literatuur blyk dat gesaghebbendes op die gebied van sielkundige toetsing toetsvolgorde as belangrik beskou, en dat daar 'n algemene neiging is om aan sekere sielkundige faktore soos aanpassing by die toetsituasie,

aan- of ontmoediging, belangstelling, ens., as die bepalende oerwegings vir die vasstelling van toetsvolgorde te dink. Hierdie oortuigings en insigte borus deurgaans op kliniese waarneming en interpretasie, en geen eksperimentele gegewens word ter staving van die (soms botsende) bewerings en aanbevelings genoem.

Daar is getuigenis dat emosionele gesteldhede soos senuagtigheid, gespannenheid, minderwaardigheidsgevoelens, ens., prestasies in sielkundige toetse nadelig beïnvloed, en dit skyn asof hierdie nadelige uitwerking opvallender is in die geval van emosioneel-wanaangepaste persone.

Dit kan ook as betreklik seker beskou word dat sekere soorte toetse gevoeliger is vir die uitwerking van afwykende emosionele gesteldhede as ander. Dit skyn nl. asof prestasies in toetse met 'n sterk visuele komponent nadeliger deur afwykende emosionele toestande getref word as wat met toetse van 'n meer verbale aard die geval is.

B. DIE RESULTATE VAN HIERDIE ONDERSOEK.

1. Hoofbevindinge.

(a) Die invloed van toetsvolgorde op toetsprestasies. Vir die bepaalde toetse in die eksperimentele batterye, en vir die bepaalde proefpersone wat in hierdie ondersoek gebruik is, is daar slegs geringe, statisties-onbeduidende verskille gevind tussen die gemiddelde toetsprestasies van die gelykwaardige groepe proefpersone, wat verskillende aanbiedingsvolgordes van dieselfde toetsbatterye ondergaan het. Die verskille was egter deurgaans in die verwagte rigting - dié groep wat die M-toets eerste gedoen het, het 'n laer M-gemiddelde as die groep wat die M-toets tweede gedoen het. (pp. 81-84)

(b) Die rol wat emosionele faktore speel. As van al die proefpersone slegs dié uitgesoek word wat 'n telling van 20 of hoër behaal het op die Persoonlike Vraelys (veronderstel om emosionele wanaangepastheid aan te dui), dan neem die verskille tussen die gemiddelde

toetsprestasies van die eksperimentele groepe toe, en kan dit in een van die twee eksperimente as statisties-beduidend beskou word. (pp. 41-66)

2. Beperkings van die eksperimentele ondersoek.

By 'n poging om die betekenis van hierdie hoofbevindinge te bepaal, moet veral gelet word op die beperkings waaraan die eksperimentele ondersoek onderhewig was, sodat die bevindinge nie op 'n ongeldige wyse veralgemeen word nie.

(a) Die bepaalde toetse wat gebruik is. Die drie potlood- en papertoesse wat gebruik is, nl. toetse AG, A/3 en M, verteenwoordig slegs 'n klein gedeelte van die groot verskeidenheid soorte toetse wat in sielkundige navorsing en praktyk gebruik word. Dit is moontlik dat met ander toetskombinasies toetsvolgordes 'n opvallender uitwerking op die toetsprestasies mag hê - hierdie aspek vereis verdere eksperimentele ondersoek. In hierdie ondersoek was daar geringe maar duidelike verskille tussen die N-gemiddeldes van die eksperimentele groepe, terwyl die verskille tussen die AG-gemiddeldes van die eksperimentele groepe, en die A/3-gemiddeldes van die eksperimentele groepe, nie konsekwent en oor die algemeen ook kleiner was. Die vernameaste verskil tussen toets M en die ander twee toetse volgens subjektiewe beoordeling is waarskynlik dat toets M 'n besondere mate van ruimtelike ordebesoef vereis, terwyl die ruimtelike komponent minder belangrik is vir die ander twee toetse. Volgens die subjektiewe beoordelings van die toetsituasies is toets M oor die algemeen, maar nie deurgaans nie, moeiliker gevind as die ander twee toetse.

(pp. 70-79)

(b) Die bepaalde proefpersone wat gebruik is. Die proefpersone wat in hierdie ondersoek gebruik is, nl. Afrikaans-sprekende Randse skoolseuns van skoolstanderds 6 tot 8, (pp. 54-60) stel ook sekere beperkings aan die veralgemeenbaarheid van die eksperimentele bevindinge. Bekendheid met skooleksamens, I.K.-toetse, en die verskillende opvoedkundige ondersoeke, wat so dikwels aan Randse skole uitgevoer word, het by hierdie

proefpersone/...

proefpersone moontlik 'n bepaalde orientasie ten opsigte van sielkundige toetsing meebring, wat heeltemal anders is as wat bv. in die militêre of 'n ander beroepskeuringsituasie aangetref word. Die proefpersone is waarskynlik daarvan bewus dat die resultate van sodanige ondersoeke hulle nie persoonlik raak nie. Buitendien bekommer hulle hulle op die jeugdige leeftyd waarskynlik min oor die persoonlike implikasies van toetsaflegging. Vergelyk mens hiermee die ouer, minder „toetswyse“, sterker gemotiveerde proefpersone, wat gewoonlik in die praktiese keuringsituasie aangetref word, blyk dit dat die bevindings van hierdie ondersoek nie noodwendig vir hulle ook sal geld nie. Die motiveringsfaktor word veral as van besondere belang beskou - as daar 'n paar begeerde vakatures is en 'n groot aantal applikante, weet elkeen in die groep applikante dat slegs 'n kwart of slegs 'n tiende van hulle aangeneem sal kan word, en dan verkry daardie toetssituasie 'n betekenis, waarmee die gewone toetssituasie (soos die onderhawige, byvoorbeeld,) waarskynlik nooit beklee word nie. Onder sulke omstandighede is die uitwerking van spanning, angs, gevoelens van sukses of mislukking, ens., waarskynlik ook opmerkliker as andersins.

Voordat die bevinding dat toetsvolgorde min verskil aan gemiddelde toetsprestasies maak veralgemeen kan word, sal soortgelyke ondersoeke met verskillende soorte sielkundige toetse, en proefpersone wat meer verteenwoordigend is van wat in die praktiese keuringsituasie aangetref word, uitgevoer moet word. Dit kan met 'n minimum moeite in enige groot keuringsprogram gedoen word, en die bevestiging van die bogenoemde bevinding sal vir die toetspraktyk waardevol wees.

(c) Die statistiese benadering van die vraagstuk. Daar moet op gewys word dat die ondersoek, en die resultate, groepe proefpersone behels, en dat die moontlikheid dat bepaalde individue kenmerke van gevoeligheid of ongevoeligheid vir toetsvolgorde sal vertoon, glad nie uitgesluit of bevestig word deur die voorgaande gevolgtrekkings nie.

3. Die Persoonlike Vraelys.

Hoewel die geldigheid van die Persoonlike Vraelys vir die meet van emosionele wanaangepastheid nie aangetoon kon word nie, toon die bevindinge dat die verklaring van die differensiële invloed van verskillende toetsvolgordes op die toetsprestasies in terme van emosionele faktore, 'n stap in die regte rigting is. Dit skyn ook asof die vermoede dat wanaangepaste proefpersone se toetsprestasies gevoeliger deur ongunstige emosionele toestande getref word, gegrond is. (Sien pp. 84-86)

Van die eksperimentele bevindinge wat die regstreeks met die doel van die ondersoek verwant is nie, is dié aangaande die betroubaarheid van skoolksamens binne dieselfde skool en standerd¹, en dié aangaande die wyse van beantwoording van die Persoonlike Vraelys waarskynlik van die belangrikste. In albei eksperimente is daar opmerklike, deurlopende en konsekwente verskille tussen die A- en B-groepe se subjektiewe beoordeling van die toetssituasie, wat waarskynlik nou verband hou met die bepaalde volgorde, waarin dié proefpersone die toetse afgelê het. Die implisiete of uitgesproke aannames in die literatuur, en ook in hierdie ondersoek, is dat die aflê van 'n toets, wat deur die proefpersoon as vreemd en moeilik ervaar word, 'n neerdrukkende uitwerking op die individu het, terwyl die aflê van 'n toets waarin die proefpersoon dink dat hy goed doen, 'n „positiewe“, „verheffende“ uitwerking het. Die A- en die B-groepe het egter dieselfde toetse, hoewel in verskillende volgordes, afgelê, en dit is 'n aansienlike veronderstelling dat die A-groepe ná voltooiing van die N-toets, 'n mening oor die N-toets gekuldig het, wat vergelykbaar was met dié van die B-groepe, /...

¹ Sien pp. 37-42. In 'n latere, nog ongepubliseerde ondersoek deur die skrywer, is gevind dat die minimum-betroubaarheid van 'n 6-vak-gemiddelde in die openbare eindexamens van die Transvaalse middelbare skole van die orde .80 is.

B-groep, direk nadat hulle die M-toets voltooi het. Dit moet dan verklaar word waarom daar sulke opvallende verskille in die subjektiewe beoordeling van die toetssituasies was (pp. 70-79), veral daar oeweging van die werklike toetsprestasies tot die teenoorgestelde gevolgtrekking as die subjektiewe beoordelings lei (pp.81-84). Een moontlike verklaring vir hierdie verskynsel is dat by die beantwoording van 'n vraelys die mees resente ervarings die grootste gewig dra, en dat die B-groep om daardie rede die M-toets deurgaans as die moeilikste toets aangedui het. Aldrie die toets wat in hierdie ondersoek gebruik is so items is minstens by benadering in stygende moeilikheidsvolgorde gerangskik, sodat die A-groep se laaste indrukke, voordat hulle met die beantwoording van die Persoonlike Vraelys begin het, die moeilikheid van die AG- (of A/3-) toets was. Die groefpersone was dus skynbaar nie instaat om hulle ervarings in die twee toetse, wat noodwendig deur 'n sekere tydsinterval geskei word, objektief met mekaar te vergelyk nie. Dit is moontlik dat hierdie verskynsel nie net tot die subjektiewe beoordeling van die toetssituasie beperk is nie, maar 'n algemene kenmerk van die vraelysmetode is - die momentele gevoelston gee die deurslag.

Hoewel die verskille nie statisties-beduidend is nie, is dit opvallend dat die B-groep deurgaans 'n hoer gemiddelde vraelyspunt as die ooreenkomstige A-groep behaal het. (pp. 86-87) As in aanmerking geneem word dat die items van die M-toets naby die einde besonder moeilik, die beginsel van matriksvoltooiing duister, en die toetstaak dus onverstaanbaar, raak, dan skyn dit asof hierdie verskil die hipotese steun dat die mees resente ervaring en die momentele gevoelston die vraelysbeantwoording bepalend beïnvloed. Die hoë betroubaarheid wat die vraelys volgens die halveringskorrelasiemetode skyn te hê (pp. 29-31) bly ook nie met hierdie beskouing oor die metingsfunksie van die Persoonlike Vraelys nie, omdat die twee helftes direk na mekaar/...

mekaar, binne 'n tydsbestek van sowat 20 minute, voltooi is. As dieselfde, of 'n parallelle vorm van die Persoonlike Vraelys na 'n langer tussenpose op dieselfde groep proefpersone toegepas sou word, sou dit waarskynlik blyk dat die skatting van die vraelys se betroubaarheid aansienlik daal, omdat dit nie net emosionele wanaangepastheid as sodanig nie, maar ook momentele gevoelston, wat met omstandighede kan wissel, moet. Omdat daar so 'n groot behoefte aan meetinstrumente op die gebied van emosionele gedraging is, en ook omdat vraelysresultate so dikwels onkrities vertolk word, is dit belangrik dat die geldigheid van die hipotese, wat hier geformuleer word, verder ondersoek word: „Beantwoording van die Persoonlike Vraelys, (en waarskynlik ook van ander aanpassingsvraelyste), word in die geval van die betrokke proefpersone (en moontlik vir alle proefpersone) nie slegs deur die meer blywende persoonlikheidsgebiede nie, maar ook deur die meer verbygaande, momentele stemmingstoestande van die betrokke proefpersoon bepaal“.

C. FINALE SAMEVATTING EN BESLUIT

1. Samevatting.

'n Eksperimentele ondersoek waarin skoolseuns as proefpersone gebruik is, en waarin die grootsmoontlike voorsorg vir nakoming van eksperimentele vereistes van konstante toetsomstandighede en gelykwaardige groepe proefpersone getref is, het aangetoon dat daar slegs geringe verskille tussen die toetsprestasies was met verskillende volgordes van toetse AG en M in die een, en toetse A/S en M in die ander eksperiment. Hoewel die verskille in die verwagte rigting was, was hulle statisties-onbeduidend en vir praktiese oorewegings waarskynlik onbelangrik. Met behulp van 'n vraelys, waarvan die metingsfunksie grootliks onbekend is, is groepe proefpersone gekies waarvoor die verskille tussen toetsgemiddeldes met verskillende

toetsvolgordes/...

toetsvolgordes groter, en in een van die twee gevalle statisties-
beduidend geword het. Aangaande die metingsfunksie van die
vraelys is gepostuleer dat dit grootliks deur momentele
stemmingstoestande beïnvloed word, en is die wenslikheid van
verdere ondersoek beklemtoon. Daar is aaruidings dat
"wanaangepaste" proefpersone swakker vaar as 'n toets wat 'n
sterk mate van ruimtelike oordeelsvermoë vereis, as 'n eerste
toets toegedien word, en dat die aflegging van só 'n toets 'n
negatiewe uitwerking op die gevoelston van die proefpersoon
het. In breë trekke dien die resultate van hierdie ondersoek
as bevestiging vir die kliniese beskouings oor toetsvolgorde
en die wisselwerking tussen toetsprestasie en emosionele
gesteldheid.

2. Besluit.

Met inagneming van die eksperimentale beperkings van
hierdie ondersoek kan soos volg besluit word:

Toetsvolgorde beïnvloed toetsprestasies slegs in 'n
geringe mate, hoewel die verskille daar, en in die
verwagte rigting, is.

Sekere emosionele faktore, tentatief as emosionele
wanaangepastheid aangedui, maak proefpersone gevoel-
iger vir die uitwerking van toetsvolgorde.

Toetse met 'n sterk ruimtelike komponent is meer
gevoelig vir die uitwerking van emosionele wanaan-
gepastheid as ander, meer verbale toetse.

Die beantwoording van 'n aanpassingsvraelys word
grootliks deur die mees resente ervarings en die
momentele stemmingstoestande beïnvloed.

BIBLIOGRAFIE

1. Ballard, P. P. The New Examiner. London: Hodder and Stoughton Limited, 1925, 269pp.
2. Baumgarten, F. Progress in Psychotechnics, 1939-1945. Bern: (Samesteller) Verlag A. Francke AG, (geen datum), 315pp.
3. Biesheuvel, S. The psychologist and selection, The Leach, 1947, Vol. 18, No. 2, pp. 14-26.
4. Burt, C. Mental and Scholastic Tests (2^e uitg.). London: Staples Press Limited, 1947, 467pp.
5. Cattell, R. B. A Guide to Mental Testing. London: University of London Press Ltd., 1936, 312pp.
6. Cattell, R. B. Description and Measurement of Personality. New York: World Book Company, 1946, 602pp.
7. Dewan, J. G. Intelligence and emotional stability, American Journal of Psychiatry, 1948, Vol. 104, pp. 548-554.
8. Du Bois, P. H. The Classification Program. Washington: U.S. (Samesteller) Government Printing Office, 1947, 394pp.
9. Fisher, V. E. An Introduction to Abnormal Psychology. New York: The MacMillan Company, 1941, 533pp.
10. Fisher, R. A. and Yates, F. Statistical Tables for Biological, Agricultural and Medical Research. (3^e uitg.). London: Oliver and Boyd, 1949, 112pp.
11. Freeman, F. S. The Theory and Practice of Psychological Testing. New York: Henry Holt and Company, 1950, 518pp.
12. Goldstein, M. J. Report on Ex-Serviceman in Sheltered Employment. Ongepubliceerde verslag in N.I.P.N.-lêer KI/28/242/10.

13. Coors, D. J. 'n Verslag oor onreëlmatighede in die beheer-
rooster vir die verdedigingsmag se siftings-
batuery. N.I.P.N.-Bulletin, 1950, Vol. 2,
pp. 4-14.
14. Coors, D. J. 'n Aantekening oor die Suid-Afrikaanse groeps-
verstandstoets. N.I.P.N.-Bulletin, 1951, Vol. 3,
No. 3, pp. 45-50.
15. Guilford, J. P. Fundamental Statistics in Psychology and Education.
New York: McGraw-Hill Book Company, Inc., 1942,
533pp.
16. Gulliksen, H. Theory of Mental Tests. New York: John Wiley
and Sons, Inc., 1950. 486pp.
17. Halstead, H. An analysis of the matrix (Progressive Matrices)
test results on 700 neurotic (military) subjects,
and a comparison with the Shipley vocabulary
test. Journal of Mental Science, 1943, Vol. 89,
pp. 202-215.
18. Hartog, P. and Rhodes, E. C. An Examination of Examinations. London:
MacMillan and Co., Limited, 1935, 81pp.
19. Hornsey, Karen. The Neurotic Personality of our Time. New York:
W. W. Norton and Company, Inc., 1937, 299pp.
20. Kendall, M. G. The Advanced Theory of Statistics, Vol. II.
London: Charles Griffin and Company Limited,
1946, 521 pp.
21. Ligen, E. M. The administration of group tests. Educational
and Psychological Measurement, 1942, Vol. 2,
pp. 337-399.

22. Moritz, J. S. The solution of normal equations by A. C. Litchen's method of pivotal condensation with an introductory note on matrix theory. N.I.P.N.-Bulletin, Vol. 1, No. 3, pp. 32-37.
23. McNemar, Q. Psychological Statistics. New York: John Wiley and Sons, Inc., 1949, 364pp.
24. Morkel, P. The A(G) Mental Alertness Tests: The reliability, difficulty and grading of the forms. 1950. Ongepubliseerde M.A.-verhandeling, Universiteit van Stellenbosch.
25. Mosier, C. I. On the validity of neurotic questionnaires. Journal of Social Psychology, 1938, Vol. 9, pp. 3-16.
26. Mursell, J. E. Psychological Testing. New York: Longmans, Green and Co., 1948, 449pp.
27. Olckers, P. J. Skoolpunte en standaarde. Tydskrif vir Maatskaplike Navorsing, 1950, Vol. 1, pp. 172-182.
28. — Psychological Abstracts, 1927-1951, Vols. 1-25, Lancaster, Pennsylvania: The American Psychological Association, Incorporated.
29. Rapaport, D., Gill, M. and Schafer, R. Diagnostic Psychological Testing, Vol. I. Chicago: New Book Publishers, Inc., 1945, 573pp.
30. Sichel, H. S. Note on reliability of combination of subtests, tests or criteria. N.I.P.N.-Bulletin, 1950, Vol. 2, No. 2, pp. 57-60.

31. Stuit, D. B. Personnel Research and Test Development in the
(Samesteller) Bureau of Naval Personnel. Princeton: Princeton
University Press, 1947, 513pp.
32. Super, D. E. Appraising Vocational Fitness by Means of
Psychological Tests. New York: Harper and
Brothers, 1949, 727pp.
33. Symonds, P. M. Diagnosing Personality and Conduct. New York:
Century, 1931, 602pp.
34. Terman, L. M. and Measuring Intelligence. London: George Harrap
Merrill, Mand A. and Company Ltd., 1949, 461pp.
35. Thorndike, R. L. Research Problems and Techniques. Washington:
Government Printing Office, 1947, 163pp.
36. Thorndike, R. L. Personnel Selection. New York: John Wiley and
Sons, Inc., 1949, 386pp.
37. Thurstone, L. L. Primary Mental Abilities. Chicago: University
of Chicago Press, 1943, 121pp.
38. Vernon, P. E. and Personnel Selection in the British Forces. London:
Parry, J. B. University of London Press Ltd., 1949, 324pp.
39. Wechsler, D. The Measurement of Adult Intelligence (2^e uitg.)
Baltimore: The Williams and Wilkins Company, 1944,
258pp.
40. Yager, J. L. The influence of emotional tension in the
intelligence test scores of children. Oprosing
in The American Psychologist, 1946, Vol. 1,
p. 464.