

PROBLEME

IN VERBAND MET VERSTANDSTOETSE

IN TWEE TALE

1. *Inleiding*: Die eerste vraag wat ons moet beantwoord, is

Wat is verstandstoetse?

Daar bestaan vandag letterlik duisende toetse. Daar is kennistoetse, aanlegstoetse, vaardigheidstoetse en dan ook dit wat as verstands- of intelligensietoetse bekend staan. Laasgenoemde is egter baie moeilik definieerbaar. Die meeste van die toetse maak daarop aanspraak dat sekere verstandelike vermoëns gemeet word. Tog bly dit moeilik om te sê wat intelligensie eintlik is. Net so is dit eintlik onwetenskaplik om op grond van toetse te sê dat een persoon meer intelligent as 'n ander persoon is, of dat een groep persone meer intelligent as 'n ander groep persone is. Ons kan alleen maar sê dat die een persoon of groep beter as die ander in *dieselfde toets* vaar. Maar dan moet dit ook dieselfde toets wees, en daarby moet teen ongewettigde veralgemening gewaak word. Prof. Manuel van Texas skrywe soos volg:

„In our work here we are avoiding as far as possible the use of the term *IQ*, and we are side-stepping the problem of the general intelligence of different linguistic and ethnic groups. Naturally, we must have names with which to label the

abilities which we measure, but the names which we give sometimes acquire meanings and associations which get us into trouble.

Intelligence is one of these. More and more, I think, we need to say “ability *A as measured by Test Y*”, instead of intelligence. The fact that English-speaking people have higher scores on Test X may be accepted, but, when we begin to interpret the score as a measure of general intellectual competence, we run into trouble — and probably we should”.

Hier het ons dus die eerste probleem in verband met verstandstoetse in twee tale, en dit is dat ons nie kan bepaal wat 'n verstandstoets presies is nie. Ons kan alleen sekere vermoëns toets, maar ons is nog nie geregtig om die resultate daarvan so te veralgemeen dat ons 'n oordeel kan uitspreek oor die verstandelike begaafdheid van persone of groepe nie. Dit geld nie alleen vir die een taalgroep teenoor die ander nie, maar ook vir die een geslag teenoor die ander. Skrywer het in sy eie studie talle toetse opgestel, en in al die toetse sonder enige uitsondering, is bevind dat seuns beter as meisies vaar. Die enigste wetenskaplike gevolgtrekking wat tot dusver hieruit gemaak kon word, was beson-

der gerusstellend: Die toetsopsteller is self van die manlike geslag. (Die toetse was omtrent almal rekentoetse). Die manlike proefpersone is blykbaar deur die aard van die toetse bevoordeel. Die tweede vraag staan nou met die eerste vraag in verband.

Hoe word 'n verstandstoets opgestel?

Die opstelling van 'n verstandstoets geskied min of meer op dieselfde wyse as met alle geÿkte toetse.

Die eerste vraagstuk wat opgelos moet word, is die doel van 'n bepaalde toets. Dit is werklik baie belangrik. 'n Geslaagde toets kan net een ding meet en dit is die ding waarvoor dit ontwerp is om juis dit te meet. As ons 'n liniaal ontwerp om meetkundige figure te meet, en ons gebruik dit as 'n termometer, dan kan ons baie interessante maar hoogs onbetroubare resultate verwag. Dieselfde sal gebeur as ons twee pasiënte se koors met 'n barometer in plaas van 'n termometer vergelyk. Altwee instrumente is van glas, het kwik en 'n lugleegte daarin, maar dit is nie die instrument wat vir die spesifieke doel ontwerp is nie. As ons die verskil in die rekenkundige aanleg van seuns en meisies wil vasstel, sal ons heeltemal anders te werk gaan, as wanneer ons 'n toets wil opstel waarin ons die rekenkundige bekwaamheid van leerlinge sonder aansien van die geslag wil meet. Presies hoe ons verskillend te werk gaan, sal straks aangetoon word. Nadat ons nou vasgestel het wat die spesifieke toets wil meet, moet ons bepaal of die toets 'n toetsbattery van subtoetse moet wees. Wanneer ons 'n vermoëtoets wil opstel wat onder die bynaam van 'n verstandstoets bekend moet staan,

dan is dit baie waarskynlik dat meer as een subtoets nodig sal wees. Verstandstoetse word soms op grond van die inhoud in twee soorte toetse onderskei, nl. verbale en nie-verbale toetse. Verbale toetse bestaan uit woorde. 'n Proefpersoon moet kan lees wat daar staan. Hy lees sekere woorde en moet sekere betrekkings tussen die woorde kan vaslê of op grond van die begrippe wat die woorde verteenwoordig of op grond van eienaardighede van die spesifieke woorde self, soos bv. die aantal letters of rym. Dit kan selfs gebeur dat 'n leerling nie besef dat hy na rymwoorde moet soek nie, 'n sekere begripsverband soek en dit vind, en volgens sy sienswyse heeltemal die regte en 'n baie vernuftige antwoord gee, maar tog geen krediet daarvoor ontvang nie. Die ergste van die saak is dat die toetsopsteller van hierdie ander „regte” antwoorde onbewus bly, en tot foutiewe gevolgtrekkings kom. Hierdie probleem word geaksentueer wanneer ons met toetse in twee verskillende tale te doen kry. Omdat die vertalings van rymwoorde selde met mekaar rym, moet van ander woorde gebruik gemaak word. Die ander woorde het ander betekenis, en die leerling of proefpersoon kan toevallig weens dieselfde oorwegings die vraag in een taal reg beantwoord en in die ander verkeerd. Veral waar ons met die sogenaamde meervoudige antwoordkeuse te doen het, word die saak baie ingewikkeld, en kan ons eintlik verwag dat „dieselfde” item in twee verskillende tale, nie meer presies dieselfde taal sal wees nie.

Die tweede groep toetse is die nie-verbale toetse. Hiervoor word hoofsaaklik van twee verskillende soorte prentjies gebruik gemaak, nl. reële en abstrakte prent-

jies. Die verband tussen verskillende prentjies moet aangetoon word.

Waar ons met reële prentjies te doen het, is die toets eintlik net gedeeltelik nie-verbaal. Waarskynlik gee die proefpersoon in sy gedagte aan elke voorstelling 'n naam, en werk in sy denke dan met die name. Daarom speel die kulturele agtergrond 'n groot rol. 'n Leerling wat dadelik die naam en gebruik van 'n artikel ken, het 'n groot voordeel bo die proefpersoon wat moet onderskei tussen die doel en funksies van „'n dinges wat so lyk” en 'n „dinges wat sus lyk”.

Dit is interessant om te verneem wat dr. Pablo Roco van Puerto Rico ('n Amerikaanse Spaanssprekende eiland in Wes-Indië) van die Lorge-Thorndike „series of non-verbal tests of intelligence” sê: „The Non-verbal series is entirely pictorial, diagramatic or numerical. The authors say that on the first two levels, it consists mostly of pictures of familiar common objects and of simple geometrical figures. What the authors do not say is that, that is true for the American children, but not for the Puerto Ricans. For example, on the very first page there appears a drawing of a *poncho* which no Puerto Rican child wil recognize as such. The *mitten* appearing on page 2 line 1 would be closer to a glove (may be a baseball glove) for the Puerto children. Even the farmer is dressed up as no farmer would in Puerto Rico . . .”

Dit is duidelik dat selfs die nie-verbale toetse waarin geen woorde voorkom nie, tog vir kinders wat verskillende tale spreek, verskillende soorte moeilikhede inhou. Ons kan net dink watter beswaar dr. Pablo Roco sou gehad het, indien daar mense in volkspeledrag in die prentjies verskyn het,

en sekere tegniese probleme in verband met jukskei, kennetjie en swartvark behandel is.

Die derde soort toets wat gewoonlik by 'n toetsbattery vir verstandsmeting ingesluit word, is rekentoetse. Gewoonlik word dit onder die nie-verbale toetse ingesluit. Oppervlakkig gesien sou ons rekentoetse kultureel seker die mees neutrale gebied kon noem. Onlangs het 'n verteenwoordiger van die departement van Bantoe-onderwys gekla oor die rekentoetse wat vir sommige proefpersone wat in hul eie taal reken, buitengewoon moeilik is. As ons sê tweehonderd-en-negentig, dan sê die Zoeloe „honderd en nog 'n honderd en dit waarby jy tien tel om nog 'n honderd te kry”. Genoeg om aan te toon dat selfs Rekenkundeprobleme in verskillende tale nie ewe moeilik is nie. Hierdie verskille kan selfs tot in die Europese tale deurgetrek word. Prof. Manuel, wat met toetse in Engels en Spaans handel, skrywe: „Why start with the assumption that one group can solve arithmetical problems just as fast and accurately as another group? Why not select a series of problems which are not culturally bound and find out what different groups can do with it — and then, of course, not overgeneralize as to the meaning of our findings?”

Omdat dit makliker is om met Arabiese syfers as met Romeinse syfers te reken, kan ons nog nie sê dat die Arabiere meer intelligent as die Romeinse was nie. Met die rekentoetse moet ons uiters versigtig wees, aangesien dit nie so 'n neutrale gebied is as wat dit op die eerste oogopslag lyk nie. Dit is bepaald kultuurgebonde.

Nadat ons nou die doel van ons toets bepaal het, en vasgestel het uit watter soort

toetse ons toetsbattery moet bestaan, is die volgende stap (die derde) om die stof vir die toets te versamel.

Wanneer ons 'n gebou optrek, word daar altyd meer materiaal aangery as wat vir die gebou self nodig is. Hoe noukeurig ons ook beplan en bereken, op die end moet daar afval-stene, sand en sement weer weggery word. Hoe beter ons beplan, hoe minder stene en ander materiaal word weer op die ou end verwyder. Gewoonlik word meer van die materiaal wat maklik verkrygbaar is op die end weggery. Op die Witwatersrand sal meer mynsand as rivier-sand weggery word. By enige gebou sal waarskynlik meer gewone stene as sier-stene na die ashoop gaan — die swakste stene die meeste.

Die toestand word nog verder verander indien besluit word om 'n skool te bou, en nadat al die materiaal aangery is, skielik besluit word om dieselfde materiaal te gebruik om 'n hospitaal te bou. Dit sal dan baie moeilik gaan om uit die afvalmateriaal af te lei of die hospitaal goed beplan is, en of die oprigting van die hospitaal met ons boukundig insigte strook. Die omstandighede waaronder die bouwerk opgerig is, maak hierdie soort oorwegings net nie meer ter sake nie.

Presies dieselfde geld vir 'n verstandstoets. Om ekonomies te werk te gaan, moet gekonsentreer word op die bymekaarbring van juis die soort items wat vir die spesiale soort toets die beste is. Gewoonlik word bevind dat twee derdes van die items bevredigend is, en 'n derde nie gebruik kan word nie. Word die toets met groter kundigheid en bekwaamheid opgestel, sal die wegrymateriaal aansienlik minder wees. Word besluit om dieselfde

items vir 'n ander soort toets te gebruik, word waarskynlik meer items uitgeskakel.

Die vierde stap is om die bymeekaargeraapte items in 'n voortoets saam te flans en om 'n aantal leerlinge of proefpersone daarmee te toets.

Van nou af word sake meer ingewikkeld en is die hulp van die statistikus daadwerklik nodig. Soos dit met die beplanning met elke projek gaan, moet die statistikus by wees wanneer die eerste lyntjie in die beplanning getrek word. Nou het ons egter die stadium bereik dat die statistikus sy moue sal moet begin oprol.

2. Die berekening van moeilikheidswaardes en diskriminasieindekse.

Die eerste vraag is nou weer: Wat is die onmiddellike doel van hierdie voortoetse?

Die antwoord is: Die berekening van die moeilikheidswaardes en die diskriminasie-indekse van die items.

In die eerste plek moet ons daarop let dat dit om *die items* gaan. Ons wil weet presies waar ons met die items staan. Waar ons met die proefpersone en met die verskillende soorte proefpersone staan, is nou nie ter sake nie. Ons wil weet hoe moeilik die item is, hoe moeilik vir seuns en hoe moeilik vir meisies. Indien die item of sy vertaling op meer as een taalgroep toegepas word, hoe moeilik dit vir taalgroep A en hoe moeilik dit vir taalgroep B is. Ons doen dit nie om taalgroep A en taalgroep B te beoordeel nie, maar om die item te beoordeel. Is dit ons doel om 'n gemeenskaplike verstandstoets vir seuns en meisies, vir stedelinge en plattelanders en vir lede van taalgroepe A en B op te stel,

dan moet ons sover moontlik op items konsentreer wat nie een van hierdie groepe ten koste van die ander begunstig nie.

Om die moeilikheid of maklikheid van 'n item te bepaal, maak ons van die sogenaamde *moeilikeidswaarde* gebruik.

a. *Die moeilikeidswaarde.*

Itemontleding kan alleen by 'n krag- of vermoëtoets toegepas word, nie 'n spoedtoets nie. Elke item moet aan die vermoë van al die betrokke proefpersone onderwerp word. Wanneer 'n item of reg of verkeerd is, en die puntetelling dus of 1 of 0, dan is die eenvoudigste moeilikeidseindeks p , die gemiddelde itempuntetelling. Ons weet wat die gemiddelde itempuntetelling vir die groep is, maar nie vir 'n bepaalde individu nie.

Ons kan hierdie getal, p , gebruik om die moeilikeidswaardes in 'n rangorde te plaas, maar sodra ons begin om dit in formules en berekenings te gebruik, moet ons versigtig wees.

In die eerste plek is dit nie 'n direkte meting van moeilikeid nie, maar van *maklikheid*. In die tweede plek verskaf dit geen lineêre metingskaal nie; dit is nie in 'n lineêre verhouding tot vermoë nie. Guilford voeg hieraan toe:

„There are two qualifications that must be mentioned. One is that such difficulty values for the same items (and their corresponding p values) will differ from one population to another, depending upon the means and standard deviations of the ability in those populations.

The other qualification is that with multiple-choice deviations or any others in which chance success is an appreciable element, a correction for chance is in order

before scaling is carried out”.

Vir dié doel stel Guilford die volgende formule ter korrigering voor:

$$cP = \frac{Kp - 1}{K - 1}$$

cP is die proporsie korrekte antwoorde (soos gekorrigeer); p is die verkreë proporsie; k is die aantal alternatiewe antwoorde.

Daar is nog ander korreksieformules wat ek nou nie almal hier wil aanhaal nie. Bostaande gaan van die standpunt uit dat alle keuse-antwoorde ewe verleidelik is, wat nie noodwendig so is nie.

Davis se moeilikeidswaarde:

Davis het 'n ander prosedure voorgestel. Hy volg die normaalkrommebeginsel soos vantevore beskrywe, maar gaan oor tot 'n lineêre transformasie van die z -skaal na 'n ander wat net positiewe waardes het, en strek van 0 tot 100 met 'n mediaan van 50. Die transformasieformule is:

$d = 21.063z + 50$, waar d as Davis se moeilikeidseindeks bekend staan. Davis het 'n tabel opgestel waarvolgens ons d direk van p kan aflei. Dit is duidelik dat Davis se tabel ons eintlik ook *maklikheidswaardes* gee. Hoe hoër die indeks is, hoe makliker is die vraag.

b. *Die diskriminasie-indeks.*

Ons wil nie net weet hoeveel persone 'n item reg beantwoord nie, maar ons wil weet hoedat die item aan sy doel beantwoord. 'n Item lewer 'n hydrae tot 'n toets, indien proefpersone wat in die toets

swak vaar en 'n lae puntetelling het, gemiddeld ook met die beantwoording van die besondere item swak vaar.

Die gewone prosedure is om die toetspersone in twee groepe te skei; 'n boonste groep en 'n onderste groep. Dit kan die boonste 50% en die onderste 50% wees; sommige verkies om die boonste 27% en die onderste 27% te neem. Dit is bewys dat wanneer met groot groepe gewerk word die 27%-metode gebruik kan word, sonder om veel aan noukeurigheid prys te gee.

Die proporsie van die boonste groep wat die item reg gekry het, word vergelyk met die proporsie van die onderste groep wat die item reg gekry het. Het 'n hoë proporsie van die „goeddoeners” die item reg gekry, en 'n lae proporsie van die „swakdoeners” dit reg gekry, dan het die item 'n hoë diskriminasie-indeks. Vaar die goeddoeners en die swakdoeners ewe goed in die item, dan is die diskriminasie-indeks nul. Vaar die swakdoeners beter as die goeddoeners, dan is die diskriminasie-indeks negatief en moet so 'n item in 'n goeie toets vermy word, aangesien dit die teenoorgestelde toets as dit waarvoor die toets opgestel is. Die afleiding van die indeks word o.a. in Guilford se „Psychometric Methods” op bladsy 425 beskrywe, en hoef nie hier herhaal te word nie. Die maklikste is om van 'n abac gebruik te maak. Dit is steeds gewens om net van items met 'n diskriminasie-indeks van ongeveer 18 en hoër volgens Davis se metode gebruik te maak. Items met 'n laer diskriminasie-indeks kan wel 'n toets langer maak, maar nie meer waardevol nie.

c. Geldigheid.

In die derde plek is dit baie belangrik

dat die geldigheid van die hele toets ook aan 'n buitekriterium onderwerp sal word, om vas te stel of die toets werklik meet wat dit voorgee om te meet, en of dit aan sy doel beantwoord en in welke mate. Kan dit in 'n besondere geval 'n voorspelling in die hand werk, of het dit net waarde as 'n „screening test?” Hierdie geldigheid moet afsonderlik bepaal word vir die verskillende groepe wat die toets bedien: seuns en meisies, platteland en stad, taal A en taal B. Vir al hierdie groepe kan 'n gesamentlike toets opgestel word, maar tog moet bepaal word watter waarde die toets vir elk het. Dit moet geskied met 'n geskikte buitekriterium as maatstaf. Die ideale toets wat vir 'n spesifieke doel opgestel is, en alle groepe met die meeste billikheid behandel, behoort ook vir alle groepe ewe groot geldigheid te hê.

3. Toetse vir verskillende groepe.

Noudat ons 'n idee het wat onder moeilikheidswaardes en diskriminasie-indekse verstaan word, kan ons met ons toetsprogram voortgaan. Hierdie waardes is nou bereken, en die regte items om in die finale toetsbattery te behou, moet uitgesoek word. Items waarvan die diskriminasie-indeks te laag is, word uitgeskakel. As die toetsopsteller gelukkig is, het hy darem nog genoeg items oor om die toets mee saam te stel. Verder rangskik die toetsopsteller die items van die maklikste tot die moeilikste en maak daarvoor van die gesamentlike moeilikheidswaardes gebruik. Die toetsopsteller het egter ook die moeilikheidswaardes van seuns en meisies apart by hom. Indien dit 'n toets is wat nie tussen hierdie twee geslagte moet onderskei nie, sal hy voorkeur gee aan items waar

die prestasie van die twee groepe naasteby gelyk is, items uitskakel waar hulle ver verskil, en daarby sorg dra dat net soveel meisiesvoorkeurigitems ingesluit word as seunsvoorkeurigitems. Daarby moet gesorg word dat die gemiddelde moeilikheidswaarde so na moontlik aan 50 is.

Die doel van die toets mag egter anders wees. Die doel mag juis wees om tussen groepe te onderskei. In daardie geval sal die toetsopsteller presies teenoorgestel te werk gaan. Hy sal die items uitsoek waarin die twee groepe juis verskillend presteer. Op die wyse kan vasgestel word nadat die toets opgestel is in watter rigting 'n persoon se aanleg lê. In 'n stelsel van gedifferensieerde onderwys kan op die manier vasgestel word na watter skool of baan 'n leerling moet gaan. Maar dan het ons heeltemal 'n ander toets as 'n gesamentlike verstandstoets. Dit kan ons 'n siftingstoets noem. Ook kan so vasgestel word watter somme beter vir meisies en watter beter vir seuns is.

Wanneer ons egter 'n gesamentlike toets vir die twee geslagte opstel, is dit veiliger om aan te neem dat die twee geslagte ewe bekwaam is, en die toets ooreenkomstig op te stel. Na die items wat weggelaat word, word eenvoudig nie weer gekyk nie. Indien die oorgeblewe items meer die seuns as die meisies sou bevoordeel het, dan weet ons dat ons volgende keer meer oordeelkundig moet wees met die versameling van ons items. Totdat die teenoorgestelde deur 'n toets wat spesiaal vir die doel opgestel is, onomstootlik bewys is, is dit veiliger om aan te neem dat die twee geslagte gelyk staan in die besondere vermoë wat die gesamentlike toets moet meet. Die groot aggregeer van items

kan nooit as 'n maat van die verskil tussen die twee geslagte aanvaar word nie, omdat dit nie deel van 'n wetenskaplike toets vorm nie, net so min as die afvalmateriaal wat na 'n bouery weggeneem word, deel van die gebou vorm. Nadat die toets opgestel is, kan vasgestel word deur toetsing aan 'n buitekriterium of die aanname reg was.

4. *Toetse in twee tale.*

Ten slotte kom ons nou by die spesifieke probleme in verband met toetse in twee tale.

Wanneer ons 'n toets vir twee verskillende taalgroepe opstel, moet ons vooraf bepaal of die twee taalgroepe gelykwaardig is in die besondere vermoë wat ons wil meet. Wil ons die verstandelike vermoë meet, is dit die veiligste weg om te aanvaar dat die vermoë gelyk is, totdat 'n behoorlike vir die doel opgestelde toetsprogram die teenoorgestelde bewys. Toevallige resultate wat met ander toetse verkry is, kan nie as 'n sodanige bewys aanvaar word nie. Nog minder kan die uitslae van 'n versameling items wat nog geen toets vorm nie, as basis en maatstaf aanvaar word. Was dit moontlik dan was toetskonstruksie en die uitsoek van items eintlik sinloos. Willekeurige saamgeraapte items kan dan sommer 'n toets vorm, sonder dat enige verder bewerking daarop toegepas word. Maar laat ons kyk wat die ondervinding in ander lande is.

Engelse en Spaanse toetse.

In Puerto Rico is die volgende toetse vertaal en kultureel aangepas:

Wechsler Intelligence Scale for Children;

Stanford-Binet, Scale Revised, Form L; Goodenough Intelligence Test.

Die Wechsler-skaal. Nadat die toets keer op keer volgens plaaslike standaarde verbeter is: „The conclusion was reached that the average Puerto Rican child scores 12 IQ points less than the average American child; in other words, IQ's of 77 to 99 as given by the Wechsler Intelligence Scale for Children adapted for Puerto Rico seem to us normal”.

Die Binet-skaal. „In conclusion, we should say that a total of 889 subjects were used in the three experiments with the Stanford-Binet scale . . . The IQ's ranged from 51 tot 134 with a mean of 95.65 and a median of 95.83 . . .” Vir Amerikaanse kinders was die gemiddelde 100.

Die Goodenough-toets. Nadat 1,767 kinders getoets is, is norme opgestel. „Comparing these norms with the norms for American children, we found that Puerto Rican children scored lower than American children”.

Beskou ons hierdie resultate oppervlakkig, dan kom ons maklik tot die gevolgtrekking dat die kinders van Puerto Rico verstandelik minder begaafd is as die Amerikaanse kinders, dat daar op die eilandjie Puerto Rico nie dieselfde prikkeling van die algemene verstandelike vermoëns plaasvind as op die Vasteland van Amerika nie. Gelukkig is nie in hierdie oppervlakkige gevolgtrekking berus nie. In 1925 het Columbia-Universiteit 'n ondersoek in Puerto Rico onderneem. In hierdie ondersoek is 69,184 toetse toegepas: 35,862 in Spaans en 33,322 in Engels. Wat was die uitslag? „Those in charge of the measurement believe that the results of the tests governing the findings of the Philippine

Survey concerning problems of equality or inequality of the intellects of races confirm the evidence that the Porto Rican people reveal as much sheer intellectual ability as do the people who now make up that great urban mass of humanity in the towns and the cities of continental United States”. Hulle kom dan tot die gevolgtrekking: „The data of the two studies argue strongly that the current techniques of test construction must recognize the linguistic factors more fully than they do now”. Dit is deur verdere toetse in die jongste tyd gestaaf.

Engelse en Franse toetse.

Ofskoon Kanada 'n tweetalige Frans-Engelse land is, speel tweetalige toetse nie 'n groot rol nie. Elke taalgroep het sy eie toetse met hul eie norms.

Die uitsondering is die Kanadese leër. Uit die aard van die saak is dit tog nodig dat hier wel 'n gemeenskaplike skaal vir die twee taalgroepe is. Anders kan dit maklik gebeur dat die een taalgroep die voorkeur kry met die uitsoek van vliëniërs en offisiere. Die offisier in bevel van hierdie aspek skrywe soos volg: „Form A of the Canadian Army's intelligence test was first developed in an English language version. Items were then translated into French, and rearranged according to difficulty values established with a French-speaking sample. The French version turned out to be much more difficult than the English version, the verbal tests yielding consistently lower scores.”

Die Kanadese leër het toe hul oplossing gesoek deur ooreenkomstig items in Frans en Engels volgens hul moeilikheidswaar-

des uit te kies, en nie volgens hul inhoud nie. Die nie-verbale items bly dieselfde vir die twee taalgroepe. In Kanada aanvaar hulle dat die intellektuele vermoë van die twee taalgroepe dieselfde is.

5. *Gevolgtrekkings.*

Ofskoon verstandstoetse met Binet in Frankryk begin het, is dit veral in die Anglo-Saksiese lande ontwikkel, en die bestaande toetse het soveel van die kultuur-agtergrond, dat persone van die kultuur gewoonlik beter as mense van ander taalgroepe soos bv. Spaans en Frans, vaar.

Waar 'n toets op vertalings gebaseer is, is die vertaalde toets gewoonlik moeiliker as die oorspronklike.

Aangesien itemontleding nie kan dien as 'n basis van vergelyking van die vermoëns van lede van verskillende taalgroepe nie, kan ook nie op grond daarvan gesê word dat die een taalgroep oor meer verstandelike vermoëns as die ander beskik nie.

Met die opstelling van twee naasteby ekwivalente toetse in twee verskillende tale, is dit wenslik om die items soveel moontlik dieselfde te hou wat die inhoud betref. Net soos wanneer toetse vir twee geslagte gesamentlik opgestel word, moet items waarvan die moeilikheidswaardes vir die twee groepe verskil uitgeskakel word, en die strewe moet wees om die moeilikheidswaardes van die twee toetse in die twee tale so na moontlik aan mekaar te bring. Wanneer aan hierdie vereistes voldoen is, kan gesamentlike norme bereken word.

Nadat die toets klaar opgestel is, moet die ondersoek voortgaan met die doel om die toetsresultate aan verdere buitekriteria te toets om sodoende die geldigheid van die toetsresultate vir die verskillende dele

van die toetsbevolking te bepaal en in die lig daarvan die toets van tyd tot tyd te hersien.

Pretoria.

A. J. van Rooy.

(Hoofvakkundige beampte, Nasionale Buro vir Opvoedkundige en Maatskaplike Navorsing).