

Tilburg University

## Het gebruik van psychologische tests bij allochtonen

van de Vijver, F.J.R.

*Published in:*  
Schakel

*Publication date:*  
1985

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*  
van de Vijver, F. J. R. (1985). Het gebruik van psychologische tests bij allochtonen. *Schakel*, 17, 29-32.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## HET GEBRUIK VAN PSYCHOLOGISCHE TESTS BIJ ALLOCHTONEN

*In dit artikel wordt een uiteenzetting gegeven over een cross-cultureel psychologisch onderzoek, dat bij de vakgroep Persoonlijheidsleer en Psychodiagnostiek op de Katholieke Hogeschool in Tilburg uitgevoerd wordt. De begeleiding van dit dissertatie-project is in handen van Prof. dr. P.J. Hettema (Persoonlijheidsleer en Psychodiagnostiek), Prof. dr. Ph. C. Stouthard (Statistiek) en Dr. Y.H. Poortinga (Cross-Culturele Psychologie).*

### De achtergrond van het onderzoek

Dit onderzoek richt zich op het ontwikkelen van psychologische tests voor allochtonen. De problemen, waarmee een constructeur van dergelijke tests te maken krijgt, zijn niet uniek. Aan docenten van klassen met zowel allochtone als autochtone leerlingen zullen een aantal van deze problemen niet vreemd zijn.

Als inleiding op het psychologisch onderzoek wil ik op enkele van deze problemen ingaan (het is onnodig te vertellen dat de hier geschetste problemen noch de enige noch de belangrijkste van deze docenten zijn).

Op de eerste plaats wil ik stilstaan bij het toekennen van beoordelingen van leerlingen in de vorm van proefwerk- of rapportpunten. Het centrale aspect waar ik hier op in wil gaan is de vraag in hoeverre beoordelingen van prestaties van allochtone leerlingen hetzelfde betekenen als beoordelingen van prestaties van autochtone leerlingen. Er bestaat een wijd scala aan opvattingen hierover. De meeste zijn te beschouwen als mengvormen van twee extremen. Aan het ene extreem is er de opvatting, die ik *universalistisch* zal noemen, waarin de prestaties van alle leerlingen onderling vergelijkbaar geacht worden. Dit betekent, dat, als een docent bijvoorbeeld een Marokkaanse en een Nederlandse leerling dezelfde beoordeling geeft, deze leerkracht een gelijk niveau van beheersing van de geëxamineerde stof bij beide leerlingen aanwezig acht. Beide leerlingen beheersen de stof even goed. De beoordelingen die binnen deze optiek gegeven worden, zijn onafhankelijk van de culturele achtergrond van de leerlingen. Een zeven is een zeven, ongeacht aan wie deze gegeven is.

Hoewel deze manier van beoordelen door velen als wenselijk gezien kan worden en het mogelijk ook op menig plaats zo gebeurt, lijkt ze niet altijd toepasbaar. Een opstel schrijven in het Nederlands stelt andere en bovenal meer eisen aan een allochtone dan aan een autochtone leerling; het zich moeten uitdrukken in een tweede taal is een moeilijkheidsfactor, die alleen bij allochtonen

speelt. (Sinds ik in Turkije gegevens voor het psychologisch onderzoek verzameld heb en mij enige beginselen van de Turkse taal eigen heb proberen te maken, kan ik me moeiteloos voorstellen, hoe moeilijk het voor een van origine Turks sprekende moet zijn om Nederlands te leren). Als een zelfde opdracht aan groepen autochtonen en allochtonen gegeven wordt kan deze qua moeilijkheidsniveau voor beide groepen zoveel verschillen, dat de docent een onderlinge vergelijkbaarheid van scores tussen deze groepen ongewenst vindt. Dit betekent dat de docent bij de beoordelingen de maatstaven per groep of zelfs per individu laat variëren (de *relativistische* opvatting). Een zeven behaald door een Marokkaanse leerling is niet hetzelfde als een zeven van een Nederlandse leerling.

Deze twee opvattingen zijn hier in extremis geschilderd. In de praktijk zullen we voornamelijk mengvormen tegenkomen. Welke specifieke mengvorm we tegenkomen zal van docent tot docent verschillen; voorts kan het ook nog eens van het ene vak tot het andere vak variëren. Zo zal het bij een vak als Nederlandse taal waarschijnlijk moeilijker zijn om de universalistische optiek te hanteren dan bij een vak als biologie.

Het geschetste continuum aan opvattingen laat zich ook vertalen als de vraag, in hoeverre een docent de culturele achtergrond van een leerling op enigerlei wijze laat meespelen in zijn/haar beoordelingen van het werk van de leerlingen.

Terloops zij vermeld dat aan beide extreme opvattingen (en hun mengvormen) nadelen verbonden zijn. Binnen de eerste, universalistische visie zal een leerling, die de Nederlandse taal niet goed beheerst, als gevolg hiervan veelal falen, ongeacht de capaciteiten van deze leerling. De kans op het verdwijnen van iedere motivatie bij deze leerling is dan reëel. De tweede, relativistische visie heeft als probleem dat het binnen een klas moeilijk te verkopen kan zijn aan de leerlingen, dat verschillende foutpatronen, c.q. aantallen fouten, tot identieke beoordelingen en vergelijkbare foutpatronen tot verschillende beoordelingen kunnen leiden. Op de omvangrijke psychologische literatuur over selectie, "test-eerlijkheid" en discriminatie zal ik hier niet ingaan.

Het tweede probleem van een docent in het intercultureel onderwijs, dat ik hier de revue wil laten passeren, heeft te maken met de inhoud van toetsen en proefwerken. De vraag, die in dit verband rijst, is aan welke voorwaarden toetsen en proefwerken moeten voldoen, willen deze een vergelijking van de prestaties van alle leerlingen, dat wil zeggen ongeacht hun culturele achtergrond, mogelijk maken. In termen van het eerdere continuüm

betekent dit dat we naar eisen vragen die we aan "universalistische tests" mogen stellen.

In mijn opinie is het van het grootste belang dat het instrument zo weinig mogelijk extra vaardigheden of kennis, anders dan de door de docent bedoelde, een beroep doet. Een voorbeeld zal verduidelijken wat ik hiermee bedoel. Een docent stelt tijdens een zeker proefwerk bestaande uit redactie-sommen de vraag, hoeveel dozijn potloden je uit vier gros potloden kunt halen. Als de begrippen "gros" en "dozijn" onderdeel vormen van de te toetsen stof, valt er weinig op het vraagstuk aan te merken. Als de docent alleen geïnteresseerd is in de vraag of de leerlingen het benodigde rekenwerk foutloos kunnen uitvoeren, kan het gebruik van de begrippen "gros" en "dozijn" een ongewenste moeilijkheidsfactor in de opgave introduceren. Een leerling, die het vraagstuk foutief oplost, kan de begrippen niet beheersen of een rekenfout gemaakt hebben. Als daarentegen een correct antwoord gegeven wordt, mogen de begrippen bekend verondersteld worden en is de benodigde rekenvaardigheid kennelijk aanwezig. Een foutief antwoord is dus moeilijker interpreteerbaar dan een goed antwoord. Interpretatie-problemen zullen groter zijn naarmate de vooraf bij de leerling bekend veronderstelde stof omvangrijker is. Het behoeft dan ook geen nader betoog dat tests of toetsen, die zo weinig mogelijk een beroep doen op allerlei andere dan de te toetsen kennis te verkiezen zijn boven tests die veel kennis veronderstellen bij de leerling. Deze laatste groep kunnen een aanzienlijke onderschatting geven van de capaciteiten van een leerling.

In de psychologie is in dit verband een onderscheid gemaakt tussen wat een persoon in aanleg kan ("competence": niet observeerbaar) en wat de persoon feitelijk doet ("performance": observeerbaar). Prestaties van leerlingen in de klas of van proefpersonen op een test behoren per definitie tot de laatste categorie. Wat een leerling aan feitelijke prestaties laat zien, hoeft echter nog geen afspiegeling te vormen van zijn of haar capaciteiten. Zo is het mogelijk, dat de leerling die niet weet hoeveel dozijn er in vier gros zit, de daartoe benodigde rekenvaardigheid wel bezit. Als het vraagstuk in een andere vorm gesteld wordt (bijvoorbeeld hoeveel groepjes van 12 potloden kun je uit 4 dozijn van 144 potloden halen? of nog simpeler: hoeveel is  $4 \times 144 : 12$ ?) weet de leerling het antwoord mogelijk wel. Als de leerling zo'n anders geformuleerde vraag beantwoordt, blijkt hij of zij de "competence" (capaciteit) wel te bezitten, maar is dit niet steeds in zijn of haar "performance" te zien. Tests, toetsen e.d. zijn beter naarmate de prestaties erop de capaciteiten van de leerling beter weergeven. In de praktijk komt dit erop neer dat de tests niet al te veel achtergrondkennis mogen veronderstellen.

De bovenstaande problemen spelen niet alleen in het onderwijs. Bij het ontwerpen van tests in de cross-culturele psychologie komen we dezelfde problemen tegen.

Het cross-cultureel onderzoeksproject dat ik nu ga beschrijven, vindt zijn oorsprong in de noodzaak om in een vanuit cultureel opzicht pluraal land als Nederland test-procedures te ontwikkelen, die voor alle culturele groepen bruikbaar zijn. In het onderzoek gaat het om tests voor inductief redeneren (abstract denken, logisch redeneren) te ontwikkelen. Deze vorm van denken is vooral van belang bij de meer exacte vakken op school. Traditioneel vormen tests voor inductief denken een essentieel onderdeel van intelligentie-tests. De eerste tests voor inductief denken dateren al van 1910.

Er bestaan binnen de psychologie enkele tests voor abstract denken, die min of meer speciaal voor cross-cultureel onderzoek ontworpen zijn. Een van de bekendste daarvan is "Raven's Progressive Matrices". Dit is een test waarbij de proefpersoon figuurreeksen moet voortzetten. Een andere is "Cattell's Culture Fair Intelligence Test". In beide tests wordt gebruik gemaakt van betekenisloze figuren; dit is gedaan om mogelijk grote culturele verschillen in bekendheid met het testmateriaal te minimaliseren. Er is veel cross-cultureel onderzoek met deze tests gedaan en telkens blijkt weer dat de inherente claims van deze tests als zouden ze "cultuurvrij" zijn (en derhalve vergelijkingen van prestaties over culturele groepen heen mogelijk maken) misplaatst zijn. Er worden vaak aanzienlijke verschillen in gemiddelde scores van een cultuur gevonden, maar er zijn sterke aanwijzingen dat deze verschillen niet zozeer te maken hebben met echte verschillen in "competence" als wel met andere factoren, bijvoorbeeld eerdere testervaring.

In ons eigen onderzoek wordt aangesloten bij de voorkeur voor betekenisloos stimulusmateriaal. Op de eerste plaats wordt binnen dit onderzoek met figuurreeksen gewerkt. Verder wordt in het onderzoek ook een test met letters gebruikt. De items hiervan bestaan uit reeksen nonsense woorden. De proefpersoon wordt gevraagd aan te geven welk groepje letters niet in het rijtje past, bijvoorbeeld CCCC JJJJ KKKK LLHT.

Ten opzichte van de bestaande tests zijn in onze testprocedure twee belangrijke wijzigingen aangebracht, die tot doel hebben de vergelijkbaarheid van scores over culturele groepen te bevorderen. Op de eerste plaats zijn de instructies, die vooraf gaan aan de test veel langer dan te doen gebruikelijk bij psychologische tests. De proefpersoon moet zo nauwkeurig mogelijk weten wat van hem of haar verlangd wordt bij het oplossen van de testitems.

In onze batterij bestaat een test uit verschillende gedeelten: eerst begint de proefleider een verbale uiteenzetting voor te lezen, hoe de test in elkaar zit en wat de proefpersoon moet doen. Daarna worden, eveneens klassikaal, een aantal voorbeelden van testopgaven gedemonstreerd door de proefleider. Als deze voorbeelden uitgelegd zijn, gaan de leerlingen zelf aan de slag. Voordat ze echter aan de oplossing van de testopgaven beginnen wordt hen eerst nog gevraagd om een aantal oefeningen te maken. Deze oefeningen worden dan (klassikaal) behandeld. Pas daarna gaan de leerlingen de testopgaven maken. Deze lange instructie, met veel herhalingen, is vooral bedoeld om allerlei verschillen tussen leerlingen, waar we niet direct in geïnteresseerd zijn en die toch van invloed kunnen zijn op de testprestatie te minimaliseren, bijvoorbeeld eerdere ervaringen met tests, een grotere kennis van de taal waarin de test afgenomen wordt, etc. (Omdat het om kunnen gaan met tijdsdruk ook zo'n factor is, worden onze tests zonder tijdsdruk afgenomen).

De tweede verandering ten opzichte van bestaande testprocedures, die in onze batterij aangebracht is, heeft te maken met de iteminhoud. Wie items van bestaande tests met figuurreeksen oplost, zal vaak getroffen zijn door het arbitraire van sommige correcte antwoorden. Het komt nogal eens voor dat de reden waarom een bepaald alternatief boven een ander verkozen is door de testconstructeur weinig overtuigend overkomt voor de gebruiker. Voor een proefpersoon die dit soort items probeert op te lossen, gaat het steeds om de vraag, wat de testsamensteller bij iedere opgave bedoeld zal hebben. Enige affiniteit van de proefpersoon met de culturele achtergrond van de constructeur zal dan ook een positieve invloed hebben op de testscore. Een groot verschil in culturele achtergrond tussen tester en geteste is dan ook een extra handicap voor de proefpersoon. Een bekende beschrijving in de cross-culturele psychologie van wat voor problemen een proefpersoon uit een niet-westerse land met westerse tests heeft, luidt: "How well can *they* do *our* tricks?".

In wat meer technische termen kunnen we dit probleem ook omschrijven als de vaagheid van het inhoudsdomen van de test voor de proefpersoon. Het is voor de proefpersoon onduidelijk wat in de ogen van de constructeur toegestane en wat niet toegestane operaties op de testfiguren zijn. Om deze situatie in onze tests tegen te gaan hebben we zowel in de figuurtest als in de test met nonsense woorden vooraf een aantal regels gedefiniëerd die de enige toegestane operaties zijn. In de woordentest is er een regel over klinkers, een regel over gelijke letters, een regel over letters na elkaar in het alfabet (bijvoorbeeld ABCD) en een regel over letters voor elkaar in het alfabet (bijvoorbeeld DCBA). Deze regels - en dit is erg belangrijk - worden tij-

dens de instructie expliciet aan de proefpersonen uitgelegd. Het inhoudsdomen van de test wordt met andere woorden volledig uitgelegd aan de leerlingen, zodat het eerder beschreven probleem van arbitrariteit waarom een bepaald antwoord het goede is, niet langer bestaat.

Het onderzoeksproject, dat ik hier beschrijf, heeft nog andere doelen dan het ontwikkelen van psychologische tests die bij uiteenlopende culturele groepen te gebruiken zijn. Een tweede doel, waar ik nader op in wil gaan, heeft te maken met de mogelijke relatie tussen kenmerken van de leeromgeving en testprestaties. Deze abstracte formulering komt erop neer, dat er onderzocht wordt of de verschillen in prestaties over culturen heen, die vaak aangetroffen worden (individueen uit cultuur A hebben gemiddeld een hogere testscore dan individuen uit cultuur B), op een of andere wijze samenhangen met verschillen in eigenschappen van het schoolsysteem. Of nog concreter gesteld: zijn er kenmerken van de school aan te wijzen, die verantwoordelijk geacht mogen worden voor de geobserveerde score-verschillen in de cross-culturele psychologie. Bijvoorbeeld, welke factoren zorgen ervoor dat, zoals vaak gevonden in Derde Wereld landen, kinderen van scholen in de stad gemiddeld hogere scores hebben dan kinderen van scholen op het platteland op de meest uiteenlopende mentale taken? Of: zijn er schoolkenmerken die zouden kunnen verklaren waarom kinderen uit de Westerse wereld gemiddeld hoger scoren dan kinderen uit Derde Wereld landen?

Om enig inzicht te krijgen in deze vraagstelling wordt er in het onderzoek een aantal kenmerken van de scholen gevraagd, bijvoorbeeld klasgrootte, aantal leerboeken, aanwezigheid van een bibliotheek, opleidingsniveau van de docenten. Daarnaast wordt er aan docenten van deze scholen een vragenlijst voorgelegd over hoe zij vinden dat een docent met zijn of haar klas moet omgaan.

Een laatste aspect van het onderzoek betreft de vraag of de ontwikkeling van abstract denken, die zich aan het eind van de lagere school en het begin van het vervolgonderwijs het meest geprononceerd voordoet, in verschillende culturen langs analoge wegen verloopt.

Al deze overwegingen hebben ertoe geleid dat we de volgende culturele groepen in het onderzoek opgenomen hebben: leerlingen uit Zambia, leerlingen uit Turkije en leerlingen uit Nederland. Deze laatste groep omvat twee subgroepen, een groep Nederlandse en een groep Turkse leerlingen.

### **De huidige stand van zaken in het onderzoek**

Er zijn op dit moment al een aantal gegevens ver-

zameld. Tijdens het voorjaar van 1983 heb ik een vijftiental scholen in Lusaka en omgeving bezocht om daar  $\pm 800$  (Zambiaanse) leerlingen te testen. In het voorjaar van 1984 zijn  $\pm 900$  leerlingen uit Istanboel en omgeving getest. Zowel in Zambia als in Turkije zijn er vier opeenvolgende klassen getest. In Zambia betrof dit de zesde en zevende klas van het basisonderwijs en de eerste twee klassen van het vervolgonderwijs. In Turkije hebben we leerlingen van de vijfde klas van de lagere school en de drie leerjaren van de middenschool getest. De Turkse leerlingen in het onderzoek waren door deze keuze gemiddeld twee jaar jonger dan hun Zambiaanse collega's. Zowel in Zambia als in Turkije werden de tests afgenomen door plaatselijke testassistenten.

Op dit moment zijn alle Zambiaanse en nagenoeg alle Turkse data geanalyseerd, zodat we een aantal voorlopige conclusies kunnen trekken. Op de eerste plaats blijkt - en dat zal geen verbazing wekken - dat zowel in Turkije als in Zambia de test scores, met name in de hogere klassen, ver boven het kansniveau liggen. De leerlingen zijn met andere woorden op de tests in staat tot abstract denken.

Verder blijken de scores met de leeftijd toe te nemen: naarmate kinderen ouder worden zijn ze in staat om meer items van de tests correct te beantwoorden. Ook dit zal weinig verbazing wekken. Verrassend is echter het feit dat de toename per klas op beide tests disproportioneel groot is tussen de zesde en zevende klas; in termen van test scores zijn de overgang van de vijfde naar de zesde en die van de zevende naar de achtste klas klein vergeleken met de overgang van de zesde naar de zevende klas. Dit verschijnsel doet zich zowel in Zambia als in Turkije voor. In deze leerjaren blijken zich dus bij veel kinderen belangrijke ontwikkelingen voor te doen op het gebied van abstract denken. Dit doet sterk denken aan analoge bevindingen in de Piagetiaanse psychologie, alhoewel in deze traditie op geheel andere wijze data verzameld worden.

Verder is ook in de Zambiaanse dataset nagegaan wat de belangrijkste variabele was ter verklaring van de gevonden score-toename: kalenderleeftijd van de proefpersoon of het aantal jaren scholing ("schoolleeftijd"). Welnu, deze laatste blijkt de belangrijkste. Als je iemands test score zou moeten voorspellen op grond van één van deze twee leeftijden zul je het verst komen met "schoolleeftijd".

Voor beide tests heb ik in beide landen hoge betrouwbaarheden gevonden (.74 - .93). De betrouwbaarheid in Zambia is meestal net iets hoger dan die in Turkije; dit komt wellicht omdat de score spreiding in Turkije kleiner is. In feite betekent dit dat het scoreverschil tussen de beste en de slechtste leerling in een klas in Zambia over het algemeen groter is dan in Turkije.

Een essentieel onderdeel van het onderzoek betrof de vraag, of het mogelijk was tests te ontwikkelen die bij groepen met een uiteenlopende culturele achtergrond te gebruiken zijn. Op basis van enkel de Zambiaanse en Turkse data is daar niet veel uitsluitel te geven. Wat ik echter wel gevonden heb zijn enige positieve aanwijzingen.

In de Zambiaanse dataset ben ik nagegaan of er veel verschil in gemiddelde scores bestond tussen leerlingen van stadsscholen en leerlingen van scholen op het platteland, zoals zo vaak aange troffen. Dit verschil blijkt miniem te zijn, hetgeen erop wijst dat op tests als de hier gebruikte de verschillen tussen stad en platteland triviaal zijn en middels een lange testinstructie en een verbeterd testdesign te reduceren of zelfs te elimineren zijn. Als deze tests in staat blijken te zijn, de prestaties van deze groepen min of meer vergelijkbaar te maken, geldt dat mogelijk ook voor scores van individuen uit verschillende culturele groepen. De kritieke toets van deze hypothese wacht echter nog op nieuwe data; om enig uitsluitel te geven moeten eerst de data in Nederland verzameld worden.

### **Wat staat er nog te gebeuren?**

Vanaf januari tot aan de zomervakantie van 1985 worden er in het kader van het project gegevens verzameld in Nederland. Aan twee groepen proefpersonen zullen daarbij de eerder beschreven tests voorgelegd worden: op de eerste plaats aan een groep Nederlandse kinderen. Het gaat hierbij om (willekeurige) vijfde en zesde klassen van het basisonderwijs en de eerste twee klassen van het vervolgonderwijs. De aanduiding "Nederlandse kinderen" zal niet helemaal de lading dekken; in deze groep zitten wel allochtone kinderen, maar het merendeel betreft autochtone kinderen.

De tweede groep die getest gaat worden, bestaat uit Turkse kinderen, die (in dezelfde leeftijdsgroepen als de Nederlandse leerlingen) in het Turks door een van origine Turkstalige psycholoog, die ook Nederlands spreekt, getest worden. Daartoe zullen in het voorjaar van '85 een aantal scholen met Turkse leerlingen om hun medewerking aan dit onderzoek verzocht worden.

In een latere fase zal ik een meer gedetailleerde weergave van de resultaten geven. Mocht u na lezing op- of aanmerkingen hebben, dan kunt u contact opnemen met Fons van de Vijver, Subfaculteit Psychologie, Postbus 90153, 5000 LE Tilburg.

F.J.R. van de Vijver