

FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN  
DEPARTEMENT PSYCHOLOGIE  
ONDERZOEKSGROEP HOGERE COGNITIE EN INDIVIDUELE VERSCHILLEN  
CENTRUM VOOR ORGANISATIE- EN PERSONEELSPSYCHOLOGIE  
TIENSESTRAAT 102 – 3000 LEUVEN



KATHOLIEKE  
UNIVERSITEIT  
LEUVEN

***Hebben autochtonen en allochtonen gelijke kansen  
bij selectieproeven met intelligentietests?***

**Samenvatting**

Dr. Michel Meulders  
Miek Vandenberk  
Prof. Dr. Paul De Boeck  
Prof. Dr. Karel De Witte  
Dr. Rianne Janssen

Maart 2005



## **Inleiding**

In het kader van het VIONA-onderzoeksproject “Psychologische testen en de effecten op de instroom van kansengroepen in het Ministerie van de Vlaamse Gemeenschap en in de Vlaamse privé-bedrijven”, werd gevraagd om voor drie kansengroepen (vrouwen, allochtonen en gehandicapten) te onderzoeken of relevante intelligentietests een bias vertonen. Deze samenvatting beschrijft het onderzoek dat gevoerd werd voor de doelgroep allochtonen. De centrale vraag van dit onderzoek is of allochtonen en autochtonen gelijke kansen hebben bij selectieproeven met intelligentietests.

## **Probleemstelling**

In de context van personeelsselectie worden scores op psychologische tests vaak gebruikt als predictoren voor bepaalde job-criteriumvariabelen zoals bijvoorbeeld succes in een bepaalde job. Hierbij wordt aangenomen dat de variabele die men beoogt te meten met de test (vb. intelligentie) empirisch predictief is voor de job-criteriumvariabele waarin men geïnteresseerd is. Het gebruik van testcores als meting van de bedoelde meetvariabele is slechts mogelijk als voldaan is aan twee voorwaarden: (1) de testcores moeten empirisch predictief zijn voor het construct dat men wil meten en (2) de testcores mogen niet door nog andere variabelen (geslacht, etnische achtergrond, ...) dan de bedoelde meetvariabele bepaald worden. Als dit toch het geval is, dan is er sprake van bias. Als de succeskans voor een item bij personen met eenzelfde intelligentie verschilt naargelang van de groep waartoe men behoort, dan spreekt men van itembias of “differential item functioning” (DIF).

In dit project worden schendingen van de tweede voorwaarde onderzocht. Meer specifiek gaan we na of allochtonen en autochtonen met gelijke vaardigheid verschillende succesansen hebben op de items van een test. Indien er verschillen worden gevonden, dan kunnen deze ook een invloed hebben op de totale testcore, zodat er discriminatie optreedt ten aanzien van een bepaalde groep.

## **Methodologie**

Voor het biasonderzoek werden testgegevens en achtergrondvariabelen van kandidaten verzameld in drie organisaties, namelijk SELOR, ABL en VDAB. Bij SELOR werden tijdens selectieprocedures gegevens verzameld van vier computertests: LOGDED (logisch redeneren), ANAVERB (verbale analogieën), CODES (code leren) en NUMVA (cijferreeksen). Bij ABL werden van alle kandidaten die in de provinciale defensiehuizen informeerden voor een functie bij het leger, gegevens verzameld van drie tests: DGEO (spatiale oriëntatie), TNV (algemene intelligentie) en WIMA (numerieke vaardigheden). Bij VDAB werd er in de lokale klantencentra bij deelnemers die een opleiding volgden, gegevens verzameld voor de Multiculturele Capaciteitentest (MCT-M). Deze test bestaat uit acht verschillende subtests: Rekenvaardigheid (getalbegrip), Componenten en Spiegelbeelden (spatiale intelligentie), Cijferreeksen en Exclusie (inductief redeneren), Kontroleren (perceptuele snelheid), Woordrelaties (woordenschat) en Woordanalogieën (verbale analogieën). Naast testgegevens werd er van elke kandidaat een

vragenlijst afgenomen waarin gevraagd werd naar de huidige nationaliteit, de nationaliteit bij de geboorte en het geboorteland van de persoon in kwestie, van zijn/haar ouders en grootouders. Op basis van de gegevens in de vragenlijst kon bepaald worden of een kandidaat allochtoon was. Hiervoor werd de definitie van VESOC gebruikt: Een kandidaat wordt als allochtoon beschouwd indien hij/zij, minstens één van zijn/haar ouders en/of minstens twee van zijn/haar grootouders bij de geboorte een niet-EU nationaliteit heeft/hebben.

Bij SELOR en ABL leverde de dataverzameling niet voldoende allochtone kandidaten op voor een betrouwbare biasanalyse en daarom werd er voor deze tests enkel onderzocht of Vlamingen en allochtonen gemiddeld verschillend presteren. Voor VDAB werden voldoende allochtone en Vlaamse kandidaten ondervraagd, zodat een volledige biasanalyse kon worden uitgevoerd. Voor elke test werd onderzocht welke items een bias vertoonden voor allochtonen versus Vlamingen. In geval van bias werd gezocht naar een inhoudelijke verklaring door de samenhang tussen de bias en bepaalde itemkenmerken te bestuderen. Vervolgens werd het effect van bias in individuele items op de testscore onderzocht. Ten slotte werd onderzocht in welke mate testcores konden verklaard worden op basis van achtergrondvariabelen zoals leeftijd, geslacht, opleidingsniveau, tijdstip van immigratie, verblijfsduur, nationaliteit van ouders, Nederlands als spreek- of moedertaal, enz.

## **Resultaten en beleidsaanbevelingen**

### *Verskil in gemiddelde prestaties van Vlamingen en allochtonen*

Het onderzoek wijst uit dat allochtonen gemiddeld lager scoren dan Vlamingen op alle subtests van de MCT-M, op de tests LOGDED, ANAVERB, CODES van SELOR en op de tests TNV en WIMA van ABL. De verschillen in gemiddelde vaardigheid, uitgedrukt in standaarddeviaties (SD) van de testcores voor Vlamingen, zijn voor de meeste subtests tussen .30 SD en .80 SD lager voor allochtonen. Voor de tests NUMVA en DGEO wordt geen significant verschil gevonden tussen allochtonen en Vlamingen. Het verschil in gemiddelde vaardigheid van Vlamingen en allochtonen is het grootst bij de verbale subtests (2 SD voor Woordrelaties en 1.5 SD voor Woordanalogieën). We stellen dus vast dat allochtonen vooral slechter presteren dan Vlamingen op de verbale subtests. Deze conclusie is gerechtvaardigd omdat alle subtests werden aangeboden aan dezelfde steekproef van Vlamingen en allochtonen. De precieze omvang van het geobserveerde verschil in gemiddeld presteren is echter niet noodzakelijk generaliseerbaar naar de volledige populatie van allochtonen en Vlamingen en naar andere onderzoeken. De steekproeven in dit onderzoek zijn immers tamelijk klein en niet representatief. Bovendien blijkt uit de literatuur dat de scores van allochtonen op verbale tests sterk bepaald worden door de verblijfsduur, het tijdstip waarop men geïmmigreerd is, en of men eerste of tweede generatie allochtoon is. Merk op dat in het onderzoek bij VDAB vooral eerste generatie allochtonen deelnamen, terwijl bij SELOR en ABL meer tweede en derde generatie allochtonen in de steekproef aanwezig waren. Dit verklaart waarom de verschillen in gemiddelde vaardigheid op de verbale tests groter zijn in het onderzoek bij VDAB.

Als allochtonen gemiddeld vooral slechter presteren op de verbale tests dan hebben zij minder kans om geselecteerd te worden voor een job en dan is er sprake van discriminatie. Deze

discriminatie is echter terecht als kennis van het Nederlands ook werkelijk essentieel is voor de succesvolle uitoefening van de job. Als het niet duidelijk is dat een betere kennis van het Nederlands leidt tot een betere jobperformantie dan is de discriminatie onterecht. In het kader van algemene wervingen (waar een duidelijk functieprofiel ontbreekt) kan men de verbale tests dus beter niet gebruiken. Indien ze toch in de testbatterij worden opgenomen, moet men ze zeker gebruiken in combinatie met andere tests.

### ***Verklaren van verschillen in gemiddelde testcores op basis van achtergrondvariabelen***

Factoranalyse van scores op de acht subtests van de MCT-M toont dat de scores van Vlamingen en allochtonen een verschillende factorstructuur hebben. In de steekproef Vlamingen kunnen alle subtestcores verklaard worden op basis van één algemene intelligentiefactor, terwijl in de steekproef allochtonen twee factoren gevonden worden: een algemene intelligentiefactor en een verbale factor. Dit betekent dat voor allochtonen de verbale subtests van de MCT-M niet alleen algemene intelligentie meten maar ook schoolse kennis en vaardigheden (meerbepaald kennis van Nederlands). Verder laten de resultaten van het onderzoek zien dat scores op de algemene intelligentiefactoren in beperkte mate samenhangen met opleidingsniveau. De scores van allochtonen op de verbale factor daarentegen kunnen iets beter verklaard worden en hangen samen met opleidingsniveau, verblijfsduur, leeftijd bij immigratie, de moedertaal en de score op een taalttest voor Nederlands. Kennis van het Nederlands, zoals gemeten met de taalttest, blijkt de belangrijkste factor om te verklaren waarom sommige allochtonen slechter scoren.

We kunnen besluiten dat voor eerste generatie allochtonen het presteren op verbale tests niet alleen afhangt van algemene intelligentie, maar ook van de mate waarin ze reeds Nederlands hebben kunnen leren. Dit hangt op zijn beurt weer af van de verblijfsduur, de leeftijd bij immigratie, de moedertaal, enz. Verbale tests zouden daarom alleen mogen gebruikt worden voor selectie als de vaardigheden die ze meten essentieel zijn voor het succesvol uitoefenen van de job. Merk op dat er bij SELOR en ABL in feite maar erg weinig eerste generatie allochtonen deelnemen aan selecties, zodat de discriminatie op basis van verbale tests in de praktijk zeer beperkt is. Indien er echter eerste generatie allochtonen deelnemen aan selecties, dan is het aangeraden om bij de interpretatie van de testresultaten de achtergrond van de kandidaat in rekening brengen. Bij kandidaten die het Nederlands gebrekkig beheersen en nog niet zolang in het land zijn, raden we aan om voor de interpretatie van de testresultaten gebruik te maken van een vergelijkbare allochtone referentiegroep. Op deze wijze wordt een meer realistische schatting verkregen van het potentieel van de kandidaat

### ***Snelheid versus vaardigheid***

De MCT-M bevat zowel zogenaamde "powertests" als "snelheidstests". Bij zuivere powertests primeert kennis en vaardigheid en zal men soms het juiste antwoord niet vinden/weten ook al heeft men zoveel tijd om na te denken als men wil. Bij zuivere snelheidstests zijn alle items ongeveer even moeilijk en kan men in principe altijd de juiste oplossing vinden, als men tijd genoeg heeft. De subtests van de MCT-M hebben allemaal een tijdslimiet maar toch nemen ze elk een verschillende positie in op het power-snelheidscontinuüm

Vlamingen presteren gemiddeld beter dan allochtonen op alle subtests (cf. supra), maar het verschil tussen Vlamingen en allochtonen in gemiddeld presteren is kleiner naarmate de test meer een pure snelheidstest is (Rekenvaardigheid, Componenten, Kontrolleren). Dit blijkt bijvoorbeeld uit de resultaten op de administratieve test Kontrolleren, waar het relatief kleine (maar significante) verschil in gemiddelde vaardigheid een puur snelheidseffect is. Anderzijds blijkt dat als kennis belangrijker wordt, het verschil in gemiddelde vaardigheid nog groter wordt. Bij de verbale tests waar voldoende kennis van Nederlands belangrijk is, zijn de verschillen in gemiddelde vaardigheid het grootst.

Hoewel allochtonen vooral minder vaardig zijn bij "powertests", zijn ze ook minder snel en werken ze minder nauwkeurig bij "snelheidstests". Bij het oplossen van de MCT-M laten allochtonen bijvoorbeeld meer items open hoewel er niet gecorrigeerd wordt voor raden bij het berekenen van testcores. Het is dus mogelijk dat allochtonen minder optimale antwoordstrategieën gebruiken en dat hun slechtere resultaten gedeeltelijk verklaard worden doordat ze minder "testslim" zijn. Om dit probleem te voorkomen, volstaat het om alle kandidaten voor te bereiden op de test. Dit kan bijvoorbeeld door op voorhand uitleg te geven over mogelijke antwoordstrategieën bij het oplossen van meerkeuzevragen, oefenopgaven aan te bieden, en uit te leggen hoe definitieve testcores berekend worden (o.a. of er een correctie voor raden wordt toegepast).

### ***DIF in de moeilijkheidsgraden van items en verklaren van DIF***

De resultaten van het onderzoek tonen aan dat bijna alle onderzochte tests een aantal "discriminerende" DIF items bevatten die relatief moeilijker zijn voor Vlamingen of voor allochtonen met dezelfde totaalscore op de test. Bij de meeste tests (Cijferreeksen, Rekenvaardigheid, Componenten, Spiegelbeelden en Kontrolleren) zijn er echter slechts weinig items met DIF en zijn er in het algemeen slechts kleine verschillen tussen succeschansen van allochtonen en Vlamingen met gelijke vaardigheid. Uitzonderingen zijn de tests Woordrelaties, Woordanalogieën en Exclusie die in 55%, 50% en 37% van de items DIF vertonen. Vooral bij de verbale tests zijn er relatief veel items waar de DIF op een beperkt deel van de schaal sterk oploopt (verschil in succeschansen groter dan .20). We kunnen besluiten dat verbale tests bij allochtonen en Vlamingen waarschijnlijk verschillende dimensies meten. Dit stemt overeen met de resultaten van factoranalyse op de subtestcores van allochtonen en Vlamingen. Bij allochtonen meten de verbale tests niet alleen algemene intelligentie maar ook kennis van het Nederlands. De scores op de verbale tests kunnen dus niet gebruikt worden om Vlamingen en allochtonen te vergelijken op vlak van algemene intelligentie.

De resultaten van onze analyses tonen dat de DIF bij verbale tests gedeeltelijk kan verklaard worden. Items zijn bijvoorbeeld relatief moeilijker voor allochtonen als ze woorden bevatten die minder vaak voorkomen in het dagelijkse taalgebruik, of als de woorden typisch zijn voor het Nederlands (de betekenis van de woorden kan minder gemakkelijk afgeleid worden op basis van een vreemde taal).

### ***Gezamenlijk effect van DIF in individuele items op de somscores***

Voor alle onderzochte tests blijkt dat de DIF-effecten in individuele items elkaar opheffen en als dusdanig geen differentieel effect hebben op de (verwachte) somscores. Na correctie voor het verschil in gemiddelde vaardigheid is het verband tussen de latente variabele en de testscore voor alle onderzochte tests hetzelfde in beide groepen. Voor deze tests kunnen de testcores dus gebruikt worden om kandidaten op een valide manier te vergelijken.

Voor tests met veel DIF (Woordanalogieën, Woordrelaties, Exclusie) zijn de somscores van allochtonen en Vlamingen echter niet vergelijkbaar omdat ze een verschillende betekenis kunnen hebben. De somscores kunnen immers het resultaat zijn van succes op verschillende verzamelingen van items die elk tegengestelde DIF vertonen.

We stellen ook vast dat bij vele tests de statistische onzekerheid in de testcores tamelijk groot is, zodat het strikt ordenen van kandidaten met een klein verschil in scores niet zinvol is. Het is dan ook aangewezen om bij het ordenen van kandidaten op basis van testcores de onzekerheid in de testcores te rapporteren, zodat onbelangrijke verschillen in testcores niet kunnen doorwegen op de besluitvorming. Om de kans op een foute rangordening te verkleinen, kan men gebruik maken van meerdere tests bij het beoordelen van een persoon, of kan men gebruik maken van een classificatiesysteem waarbij personen in een bepaald scorebereik worden toegewezen aan dezelfde klasse.

### ***Verdere aanbevelingen***

Om een gelijke kansenbeleid te voeren dient men te vermijden dat kansengroepen ten onrechte gediscrimineerd worden op basis van tests. Dit kan door de kwaliteit van tests op een aantal punten systematisch te bewaken. Zoals bij geneesmiddelen en voedingswaren al het geval is, zou men de kwaliteitscontrole op tests die gebruikt worden in selecties bij de overheid en in de privé-ondernemingen kunnen verscherpen en systematiseren. Als een nieuwe test op de markt gebracht wordt door een testproducent, zou men de test een kwaliteitslabel kunnen toekennen dat aangeeft in welke mate de test onderzocht is op aspecten als betrouwbaarheid, validiteit, afwezigheid van discriminatie ten aanzien van kansengroepen enz. Het toekennen van een kwaliteitslabel zou bijvoorbeeld kunnen gebeuren door een onafhankelijke commissie van testexperts (zoals bijvoorbeeld de Commissie Psychodiagnostiek van de Belgische Federatie van Psychologen).

Om in de toekomst de kwaliteit van tests (bij de overheid) te bewaken en te optimaliseren dient men de ruwe testgegevens en relevante achtergrondvariabelen na de selectie te inventariseren. Vervolgens kan men op basis van de analyse van deze gegevens de kwaliteit van de tests optimaliseren. Deze optimalisatie zou bijvoorbeeld kunnen gebeuren door testexperts in dienst van de overheid, of zou bijvoorbeeld kunnen uitbesteed worden aan een wetenschappelijke instelling.

Hoe een goede en betrouwbare rangordening te maken van kandidaten op basis van de scores die behaald werden op basis van één of meerdere selectietests blijkt een belangrijk vraagstuk te zijn dat ook ter sprake komt in het eerste en tweede deelrapport. In de besprekingen in verband met deze onderzoeken blijkt dat een afstemming op het terrein tussen enerzijds de verwachtingen van

de overheid als opdrachtgever en anderzijds de selecteurs, rekening houdende met de aanbevelingen in de deelrapporten, zich opdringt.