



de Rechtspraak

Gezocht: Allochtone rechters

Een analyse van de cognitief/analytische test uit het Raio-selectieproces

Annie Reitsma

Hengelo, juni 2005

Gezocht: Allochtone Rechters

Een analyse van de cognitief/analytische test uit het Raio-selectieproces

Annie Reitsma (s0055468)

Bachelorthese in het kader van de opleiding Arbeids- en Organisationspsychologie aan de Universiteit Twente

Universiteit Twente, Enschede
Faculteit Gedragwetenschappen

Raad voor de Rechtspraak
Juliana van Stolberglaan 4
2509 LP Den Haag

Begeleiders:

Dr. R.R. Meijer

Dr. Ir. B.P. Veldkamp

Drs. E.R. van Winzum (externe begeleider)

Dhr. M. Groeneveld (externe begeleider)

Samenvatting

In dit verslag zal het onderzoek naar mogelijke discriminatoire effecten in de Raio-selectieprocedure uiteen gezet worden. Deze selectieprocedure bestaat uit vijf fasen:

- Toetsing van formele vereisten
- Cognitief/analytische test
- Voorgesprek
- Assessment
- Commissiegesprek of selectiegesprek

Er zal in dit onderzoek alleen gekeken worden of er discriminatoire effecten aanwezig zijn in de cognitief/analytische test, dit omdat dit de belangrijkste fase van het selectieproces is, waar de meeste kandidaten afvallen.

Dit onderzoek wordt gedaan op verzoek van de Raad voor de Rechtspraak te Den Haag en dient ter afronding van de Bacheloropleiding psychologie aan de Universiteit Twente te Enschede. De Raad voor de Rechtspraak kwam met het verzoek voor dit onderzoek, omdat zij het gevoel heeft dat er bij de cognitief/analytische test meer allochtone dan autochtone kandidaten afvallen. Om dit te onderzoeken staat de volgende onderzoeksvraag centraal: *“Zijn er onbedoelde discriminatoire effecten in de cognitief/analytische test tussen allochtonen en autochtonen?”* Gezien de beschikbare gegevens is het echter niet mogelijk om te kijken of de cognitief/analytische test itembias of testbias heeft. Om deze reden zal er naar enkele andere aspecten gekeken worden, welke kunnen duiden op een onbedoeld onderscheid in de test. Deze aspecten zijn: mogelijke verschillen in subtestscores en het verschil in somscore tussen allochtonen en autochtonen. Ook zal er in dit onderzoek gekeken worden naar verschillen tussen allochtonen die de uitgebreide versie van de test doen en allochtonen die de standaardversie van de test doen. Als laatste wordt er gekeken of er een trend te zien is in de testscores over de jaren heen. Dit houdt in dat er naar de mogelijkheid gekeken wordt, dat de allochtonen nu hoger scoren dan enkele jaren geleden. Aansluitend op deze aspecten zullen dit onderzoek enkele subhypothesen centraal staan, namelijk de volgende:

1. *Er wordt verwacht dat allochtonen die de uitgebreide versie van de cognitief/analytische test maken een grotere kans van slagen hebben voor deze fase van het selectieproces;*
2. *Er wordt verwacht dat er een verschil is in de scores tussen de verschillende subtests;*
 - a. *Er wordt verwacht dat de score op de DAT '83 (zinnen) significant lager ligt dan het gemiddelde op de andere testscores;*
 - b. *Ondanks dit wordt er verwacht dat de score op de LTP-verbaal redeneren test lager zal liggen dan de score op de andere redeneer test (LTP-abstract redeneren);*
3. *Er wordt verwacht dat er een vooruitgang te zien is in de testscores over de jaren heen.*

Voordat er antwoord wordt gegeven op de onderzoeksvraag en de hypothesen getoetst worden, heeft er een literatuurstudie plaatsgevonden om veel van de gepubliceerde literatuur over het testen van allochtonen samen te vatten en een duidelijk beeld te krijgen met betrekking op de problematiek betreffende het testen van allochtonen. In de reeds gepubliceerde onderzoeken wordt veel gezegd over het testen van allochtonen. De een zegt dat het geen problemen oplevert en dat allochtonen de reeds ontwikkelde testen met gemak kunnen maken. Anderen zeggen dat dit niet mogelijk is en dat er voor het testen van allochtonen wel degelijk maatregelen genomen dienen te worden.

In dit verslag wordt gekeken in hoeverre er ondersteuning te vinden is voor deze standpunten. Dit wordt gedaan door middel van het verkrijgen van kwalitatieve en kwantitatieve data. De kwalitatieve data betreffen de subtestscore en de somscore behaald door de allochtone kandidaten welke deelgenomen hebben aan het selectieproces vanaf 2000 tot nu.

De kwantitatieve data zijn verkregen bij het LTP, door middel van een interview en observatie van een gesprek tussen de psycholoog en een solliciterende allochtone kandidaat. Door middel van data-analyse zijn de hypothesen getoetst.

Uitkomsten van het onderzoek zijn:

- dat er geen significant verschil is tussen allochtonen welke de standaardversie van de cognitief/analytische test maken en allochtonen welke de uitgebreide versie van de test maken. In beide gevallen gaat ongeveer eenzelfde percentage door naar de volgende fase van het selectieproces;
- dat er een significant verschil in scores is tussen de verbale redeneertest en de abstracte redeneertest. De allochtone kandidaat scoort aanzienlijk hoger op de LTP-abstract redeneren;
- dat er geen significant verschil is tussen de DAT '83 en de overige testcores en dat er eveneens geen significant verschil gevonden is tussen de overige subtestcores vergeleken met de andere scores;
- dat er geen significant vooruitgang is te zien in de testcores van de allochtone kandidaten genomen over de afgelopen vijf jaar.

Helaas kan er niet met zekerheid een antwoord op de hoofdvraag gegeven worden, dit komt door een gebrek aan gegevens. Wel is de kans groot dat er door de test een onderscheid gemaakt wordt tussen allochtonen en autochtonen, dit komt uit de resultaten naar voren. Om deze reden is de belangrijkste aanbeveling voor de Raad voor de Rechtspraak om een uitgebreider vervolgonderzoek uit te voeren, waarbij men zich in het bijzonder richt op de itemscores van de tests. Op deze manier kan er met meer zekerheid een antwoord gegeven worden op de gestelde hoofdvraag.

Inhoudsopgave

Voorwoord	5
1. Inleiding	6
1.1 Verantwoording	6
1.2 Aanleiding voor het onderzoek	7
1.3 Probleemstelling	8
2. Raad voor de Rechtspraak	10
2.1 Het selectieproces	10
2.1.1 Formele vereisten	10
2.1.2 Cognitieve/analytische test	11
2.1.3 Voorgesprek	12
2.1.4 Assessment	12
2.1.5 Commissiegesprek of selectiegesprek	14
2.2 Raio-opleiding	14
3. Testgebruik bij allochtonen	16
3.1 Algemeen	16
3.2 Cognitief/analytische test	17
3.3 Bias	19
3.4 Kritische kanttekeningen	20
3.5 Conclusies	21
4. Methoden	22
4.1 Testgegevens	22
4.2 Procedure	23
4.3 Verantwoording aanpak	24
5. Resultaten	25
5.1 Beschrijving data	25
5.2 Analyse	26
5.3 Samenvatting resultaten	28
6. Discussie	29
6.1 Resultaten	29
6.2 Selectieproces	30
7. Aanbevelingen	31
7.1 Tekortkomingen en aanbevelingen	31
7.2 Vervolgstappen	31
Referenties	33

Voorwoord

Met het afronden van dit verslag is ook het einde van mijn bachelor psychologie in zicht. Ik kijk met veel plezier terug naar mijn studententijd in Enschede. Het is moeilijk om te geloven dat deze zo meteen afgelopen is. Nu is het zover dat ik mijzelf Bachelor of Science mag noemen. Ik dacht dat vooral het laatste jaar van mijn studie en het afstuderen in het bijzonder erg zwaar zouden zijn, maar het afgelopen jaar was erg leuk interessant en leerzaam waardoor ik mij er voor de volle honderd procent heb ingezet.

Natuurlijk ben ik niet alleen zover gekomen. Graag wil ik daarom enkele mensen bedanken die mij hebben geholpen. Allereerst wil ik mijn ouders bedanken. Dat zij mij steunden in de keuzes die ik heb gemaakt. Ook wil ik Maarten bedanken, hij heeft me er doorheen geholpen als ik door de bomen het bos niet meer zag, vooral de afgelopen paar weken.

Ook wil ik de Raad voor de Rechtspraak bedanken, in het bijzonder Elko van Winzum en Martin Groeneveld. Zij hebben de benodigde informatie verstrekt en hebben het verkrijgen van de data mogelijk gemaakt. Natuurlijk bedank ik ook Michiel Tillema hartelijk voor wat hij allemaal gedaan heeft. Zonder hem was dit verslag niet tot stand gekomen. Hij heeft zijn best gedaan om alle benodigde data boven tafel te krijgen.

Last but not least ben ik ook mijn begeleiders, Rob Meijer en Bernard Veldkamp erg dankbaar voor het ondersteunen en de nuttige feedback, maar vooral dat duwtje in de goede richting.

De afgelopen periode, met het doen van het onderzoek en het schrijven van dit verslag, was zeer leerzaam. Ik kijk er met veel plezier op terug.

Hengelo, juni 2005
Annie Reitsma

1. Inleiding

Dit onderzoek wordt uitgevoerd in opdracht van de Raad voor de Rechtspraak. Zij heeft aangegeven dat zij meer allochtonen voor de Raio-opleiding (opleiding tot rechterlijke ambtenaar) zouden willen hebben. Er is namelijk aangegeven dat er erg veel allochtonen uitvallen tijdens het selectieproces, vooral als dit vergeleken wordt met de uitval van autochtonen. De Raad voor de Rechtspraak streeft naar meer allochtone rechters, zodat de Rechterlijke Macht een meer representatieve verdeling van de bevolking krijgt. Dit houdt in dat men meer allochtone rechters in dienst heeft of krijgt, dan nu het geval is, zodat de gerechtelijke samenstelling meer representatief, maar niet volledig representatief, is voor de Nederlandse bevolking.

In dit verslag zal er gekeken worden naar deze problematiek. Enerzijds zal er een literatuurstudie plaatsvinden welke de problematiek van het testen van allochtonen beschrijft. De vele artikelen die geschreven zijn en de vele onderzoeken die gedaan zijn zullen worden samengevat. Anderzijds zal er een analyse van de selectiedata van de cognitief/analytische test plaatsvinden. Uiteindelijk zal een advies uitgebracht worden met betrekking op het selectieproces specifiek gericht op de cognitief/analytische test. In de rest van dit hoofdstuk zal de aanleiding van het onderzoek en de probleemstelling besproken worden.

Hierna zal in de komende hoofdstukken getracht worden een antwoord te geven op de gestelde onderzoeksvraag en de opgestelde hypothesen getoetst worden, welke verderop in dit hoofdstuk aan bod komen. In de komende hoofdstukken zal eerst informatie over de Raad voor de Rechtspraak gegeven worden, waarin het selectieproces beschreven worden, met de daarbij behorende verschillende fasen, en waar ook kort de Raio-opleiding beschreven zal worden. Na deze informatie zal de recente en relevante literatuur besproken worden over het testen van allochtonen en de bijbehorende problematiek. Hierna zullen de gehanteerde methoden voor het onderzoek besproken worden. Hierbij wordt ingegaan op het verkrijgen van zowel de kwalitatieve als de kwantitatieve data, welke gebruikt worden voor het beantwoorden van de onderzoeksvraag. Als de onderzoeksmethoden besproken zijn, zal overgegaan worden op de data-analyse en zullen de gevonden resultaten besproken worden. Aan de hand van deze resultaten zal de probleemstelling beantwoord worden en zullen de conclusies van het onderzoek gegeven worden. Als dit gedaan is zullen de discussiepunten van het onderzoek besproken worden. Het geheel zal afgesloten worden met enkele punten ter aanbeveling voor de Raad voor de Rechtspraak.

1.1 Verantwoording

Voordat over wordt gegaan naar de kern van het onderzoek dient de betekenis van het woord allochtoon verduidelijkt te worden. Het woord allochtoon wordt erg veel gebruikt de laatste tijd, maar er wordt niet altijd hetzelfde bedoeld. Met het woord allochtoon worden vaak niet-westerse allochtonen of alleen Turken, Marokkanen, Surinamers en Antillianen (TMSA's) bedoeld (Intelligence Group, 2004).

In dit onderzoek zal er gekeken worden naar verschillen tussen autochtonen en allochtonen in het selectieproces waarbij specifiek gekeken wordt naar de verschillen in testscores op de cognitief/analytische test. In veel artikelen wordt geen definitie gegeven van het begrip allochtoon of etnische minderheid.

Om onduidelijkheden te voorkomen en omdat er geen duidelijke definitie gehanteerd wordt door de Raad, zal dat hier echter wel gedaan worden. In de rest van dit verslag zullen de term allochtoon en etnische minderheid door elkaar gebruikt worden, maar wordt hetzelfde bedoeld. In dit verslag zal iedereen die "Nederlands niet als moedertaal heeft", beschouwd worden als allochtoon. Deze definitie is specifiek maar komt enigszins overeen met de definitie die het Centraal Bureau voor de Statistiek (2005) van allochtonen geeft: "Een persoon van wie ten minste één ouder in het buitenland is geboren". Belangrijk is dat men zich realiseert dat het hier gaat om zowel Westerse als niet-westerse allochtonen en dat het niet alleen gaat om Turken, Marokkanen, Surinamers en Antillianen (TMSA's) zoals vaak gehanteerd wordt.

1.2 Aanleiding voor het onderzoek

Momenteel zijn er aanzienlijk minder allochtone rechters werkzaam bij de Rechterlijke Macht dan de Raad voor de Rechtspraak graag zou willen. Tevens zijn er veel minder allochtone dan autochtone kandidaten die solliciteren voor de Raio-opleiding (vergelijk tabel 1.1 en tabel 1.2 voor de instroomgegevens). Rustig afwachten totdat dit vanzelf aantrekt en er meer allochtonen bij de Rechtspraak werkzaam worden, wil men niet. Men wil graag dat er meer allochtone rechters in dienst komen. Om te kijken waar het aan ligt dat er minder allochtone rechters zijn, wil de Raad kijken of er directe oorzaken zijn voor de scheve verdeling van de rechters. Als men deze oorzaken kent, kan er gekeken worden welke men helemaal of gedeeltelijk op kan lossen.

De scheve verdeling kan liggen aan het wervingsbeleid van de Raad voor de Raio-opleiding. Hier zal naar gekeken worden door Miltenburg (2005), wie, gelijktijdig met dit onderzoek, onderzoek naar het wervingsbeleid doet. De scheve verdeling kan ook aan het selectieproces liggen voorafgaand aan de opleiding. Mogelijk wordt er in het selectieproces onbedoeld een onderscheid gemaakt tussen autochtonen en allochtonen. Bijvoorbeeld doordat de gebruikte tests niet goed bruikbaar blijken te zijn voor allochtonen. Een van de redenen hiervoor kan zijn dat de allochtonen de test of de testinstructies niet goed begrijpen.

Tabel 1.1. Instroom allochtonen/autochtonen voor de Raio-opleiding (Getallen zijn inclusief allochtonen)¹

T.b.v. de opleiding	Totaal aantal	Geslaagd	Geslaagd	Geslaagd	Negatief toch	Naar de opleiding:
april en oktober aanmeldingen: analytische test: voorgesprek: assessment: sel.gesprek:						
2000 ²	127	60	49	40	7	26
2001	252	128	104	82	16	56
2002	232	118	94	79	11	52
2003	259	131	112	103	7	65
2004	357	195	139	106	18	55
2005 ³	178	99	75	46	8	24
Totaal	1405	731	573	456	67	278

¹ Uitgegaan is van de aantallen per opleiding. Per jaar starten 2 opleidingen: april en oktober.

² Betreft opleiding oktober 2002

³ Betreft opleiding april 2005

Tabel 1.2. Instroom allochtonen voor de Raio-opleiding

T.b.v. de opleiding	Totaal aan- meldingen:	Aantal allochtonen:	Geslaagd analytische test:	Geslaagd voorgesprek:	Geslaagd assessment:	Negatief toch sel. gesprek:	Naar de opleiding:
2000 ²	127	6	1	1	1	0	1
2001	252	14	3	2	1	1	2
2002	232	15	5	4	2	1	2
2003	259	12	4	4	3	1	4
2004	357	21	3	1	1	0	1
2005 ³	178	19	3	2	1	0	1
Totaal	1405	87	19	14	9	3	11

1.3 Probleemstelling

In dit verslag zal er gekeken worden of er in het selectieproces onbedoelde discriminatoire effecten optreden. Dit houdt in dat er mogelijk een onbedoeld onderscheid gemaakt wordt tussen autochtonen en allochtonen. In dit onderzoek zal de nadruk liggen op de cognitief/analytische test (CAT). Hier wordt de nadruk op gelegd, omdat uit gegevens van de Raad voor de Rechtspraak blijkt dat de uitval van allochtonen vergeleken met de uitval van autochtonen bij deze test erg groot is, maar ook omdat deze test een zeer belangrijke stap is in het selectieproces. Om te kijken of er daadwerkelijk effecten zijn zullen de gegevens statistisch geanalyseerd worden. Er zal gekeken worden naar de verdeling van de scores. Daarnaast zullen de verschillende subtests geanalyseerd worden.

Om op bovenstaande een antwoord te krijgen zal getracht worden om op de volgende vraag een antwoord te krijgen:

Zijn er onbedoelde discriminatoire effecten in de cognitief/analytische test tussen allochtonen en autochtonen?

Bij deze vraag moet echter vermeld worden dat het niet mogelijk is om een direct antwoord op deze vraag te geven. Gezien de beschikbare gegevens is het niet mogelijk om te onderzoeken of er itembias of testbias aanwezig is in de test. Voor het onderzoeken van itembias zijn itemscores nodig en voor testbias zijn predictor- en criteriumwaarden noodzakelijk. Helaas is het niet mogelijk gebleken om deze scores te verkrijgen. De vraag zal zo goed mogelijk beantwoord worden door enkele andere aspecten, welke met onbedoelde effecten te maken hebben, te onderzoeken. Deze aspecten zijn, het verschil in scores bij de standaardversie van de test en de uitgebreide versie van de test, verschillen in scores tussen de subtests en de ontwikkeling van de test scores over de jaren.

Aansluitend op de vraagstelling zijn er enkele verwachtingen, te weten:

1. *Er wordt verwacht dat allochtonen die de uitgebreide versie van de cognitief/analytische test maken een grotere kans van slagen hebben voor deze fase van het selectieproces.* Dit wordt verwacht omdat hier in de literatuur veel over gesproken is. Door een uitgebreidere test en meer tijd, daar waar mogelijk en benodigd, is er meer tijd voor de personen om zich te bewijzen en te laten zien dat ze voldoende kennis bezitten om door te mogen gaan naar de volgende fase (Van den Berg & Van Leest, 1999).
2. *Er wordt verwacht dat er een verschil is in de scores tussen de verschillende subtests.* Dit houdt in dat verwacht wordt dat de kans op een negatieve score op de ene subtest groter is dan een negatieve score op een andere. Hierop aansluitend kunnen de volgende subhypotheses gesteld worden:
 - a. *Er wordt verwacht dat de score op de DAT '83 (zinnen) significant lager ligt dan het gemiddelde op de andere testscores.* Dit wordt verwacht omdat de DAT '83 een test is waar men direct wordt getest op taalkennis. Bij de andere tests gaat het meer om redeneren (Van den Berg & van Leest, 1999).
 - b. *Ondanks dit wordt er verwacht dat de score op de LTP-verbaal redeneren test lager zal liggen dan de score op de andere redeneer test (LTP-abstract redeneren).* Dit omdat het bij de verbale redeneertest zeer belangrijk is dat men weet wat de woorden precies betekenen. Mocht de talenkennis echter niet helemaal goed zijn, zal dit ook te zien zijn in de testscores van deze test. Een uitgebreide beschrijving van de tests is terug te vinden in het volgende hoofdstuk.
3. *Er wordt verwacht dat er een vooruitgang te zien is in de testscores over de jaren heen.* Hiermee wordt bedoeld dat de kans groot is dat de allochtonen zich de laatste jaren goed ontwikkeld hebben waardoor zij in 2005 hoger scores op de test dan in de voorgaande jaren (Te Nijenhuis, De Jong, Evers & Van der Flier, 2004).

In de volgende hoofdstukken zal de onderzoeksvraag beantwoord worden en zullen de gestelde hypothesen getoetst worden. Maar voordat dit gedaan wordt, zullen eerst enkele recente publicaties over het testen van allochtonen beschreven worden.

2. Raad voor de Rechtspraak

Op 1 januari 2002 werd de Raad voor de Rechtspraak opgericht. Dit gebeurde gelijktijdig met de invoering van een integraal management binnen alle gerechten. Deze veranderingen werden ingevoerd om de bedrijfsvoering te verbeteren en om slagvaardiger en moderner te worden en te kunnen reageren (Raad voor de Rechtspraak, 2005).

De Raad voor de Rechtspraak vormt de schakel tussen de Minister van Justitie en de gerechten. De Raad zorgt ervoor dat de gerechten hun taak goed vervullen. Tevens behartigt de Raad de gezamenlijke belangen van de gerechten, om zo een duidelijk beeld voor de buitenwereld te vormen. Hiernaast draagt de Raad voor de Rechtspraak ook zorg voor gerechtsoverstijgende voorzieningen, houdt het toezicht op de bedrijfsvoering en financieel beheer en geeft, mocht het nodig zijn, aanwijzingen met betrekking tot de bedrijfsvorming. Door deze vele taken en situaties vervult de Raad een coördinerende, stimulerende, faciliterende, toezichthoudende of sturende rol.

2.1 Het selectieproces

Er zijn twee wegen naar de aanstelling als rechter. Dit kan via de Raio-opleiding (Rechterlijke Ambtenaar In Opleiding) en de Rio-opleiding (Rechter In Opleiding). Voor beide opleidingen zijn toelatingscommissies. Voor de Raio-opleiding is dit de Raio-commissie en voor de Rio-opleiding is dit de CALRM (de Commissie Aantrekken Leden Rechterlijke Macht). In de rest van dit verslag zullen de CALRM en de Rio-opleiding volledig buiten beschouwing worden gelaten. Onderstaand zal verder ingegaan worden op het selectieproces welke de kandidaten goed dienen te doorlopen voordat zij toegelaten kunnen worden tot de Raio-opleiding.

Het selectieproces zoals dat momenteel gehanteerd wordt voor de Raio-opleiding is samen met het LTP in 1999 opgezet. Dit omdat de Raio-commissie ontevreden was met het selectieproces zoals dat toentertijd verliep in samenwerking met de Rijkspsychologische Dienst (RPD). Er is toen gekozen om samen met een ander psychologisch adviesbureau, het LTP, een nieuw selectieproces op te zetten. Vanaf januari 2000 ging het nieuwe selectieproces officieel van start, maar pas rond april 2000 kwamen de eerste kandidaten voor de nieuwe selectieprocedure. De doorlooptijd van het proces is tegelijkertijd met het opzetten van de nieuwe procedure teruggebracht van 8 maanden naar 6 tot 8 weken.

Het selectieproces bestaat uit verschillende onderdelen of fasen. Deze fasen dient de kandidaat allemaal goed te doorlopen voordat hij of zij kan beginnen met de opleiding. In totaal is het proces op te splitsen naar vijf onderdelen of beoordelingsfasen. Deze onderdelen zullen hieronder één voor één beschreven worden. De informatie over het selectieproces komt veelal uit het verslag van Cohen (2003). Daarnaast is er gebruik gemaakt van informatie aangereikt door de Raad voor de Rechtspraak (2005).

2.1.1 Formele vereisten

Voor het verzoek tot toelating tot de Raio-opleiding dient de sollicitant sollicitatieformulieren in te vullen. In de eerste fase wordt gecontroleerd of de sollicitant voldoet aan de formele ingangseisen. Zo moet men de Nederlandse nationaliteit bezitten en moet men de studie Rechten hebben afgerond.

Hierbij dient men de volgende onderdelen gevolgd te hebben: Nederlands burgerlijk (proces)recht, straf(proces)recht en bestuurs(proces)recht. Men kan zich ook voor het afronden van de Rechtenstudie aanmelden, maar dan dient de afstudeerdatum vóór de aanvangsdatum van de opleiding te liggen (april of oktober).

Toetsing van bovengenoemde voorwaarden wordt door het secretariaat van de Raio-selectiecommissie gedaan. Als men niet voldoet aan de formele vereisten of als de sollicitatie er onverzorgd uit ziet, wordt de kandidaat afgewezen. Nadat een kandidaat is afgewezen, het maakt niet uit in welke fase, kan hij na drie jaar nogmaals solliciteren voor de opleiding. Bij een tweede afwijzing is dit niet meer mogelijk.

2.1.2 Cognitieve/analytische test

Als de kandidaat voldoet aan de formele vereisten, wordt men toegelaten tot de volgende fase van het selectieproces. Hier wordt de kandidaat getoetst op intelligentie door middel van de cognitief/analytische test. Bij de cognitief/analytische test wordt men op verschillende onderdelen gescoord. Bij een voldoende score op de gehele test (men hoeft niet op alle vier testen een voldoende te behalen, er kan ook gecompenseerd worden), wordt men toegelaten tot de volgende fase van het selectieproces. Als men onvoldoende scoort, mag er niet deelgenomen worden aan de overige fasen van het selectieproces.

De test bestaat uit vier onderdelen: de LTP-abstract redeneren, de LTP-verbaal redeneren, de differentiële aanleg test '83 en een syllogistentest. Allochtonen kunnen voor de cognitief/analytische test een gesprek met een psycholoog van het LTP aanvragen. Dit kan gedaan worden om te kijken of er behoefte is aan een uitgebreide versie van de test. Mocht hier behoefte aan zijn, krijgt men een meer uitgebreide versie van de cognitief/analytische test. Bij deze versie van de test krijgt men onder andere de Woordenlijst Allochtonen voorgelegd. Hieronder staat een uitgebreide uitleg van de verschillende testonderdelen.

Woordenlijst allochtonen (WLA)

Deze test is door het LTP ontwikkeld en wordt voorgelegd aan de allochtonen die er voor kiezen om de uitgebreide versie van de cognitief/analytische test te doen. In deze lijst wordt gevraagd om de betekenis te geven van alledaagse en veel voorkomende en gebruikte Nederlandse woorden en enkele gezegdes. Voor Nederlanders is dit zeer makkelijk, maar voor veel allochtonen blijkt dit vaak moeilijker te zijn dan men verwacht. In totaal bevat de WLA 26 items. Als er teveel fouten gemaakt worden, heeft de allochtone kandidaat niet voldoende kennis van de Nederlandse taal om de persoonlijkheidsvragenlijsten uit een later stadium van de selectieprocedure goed te begrijpen en goed te kunnen invullen.

LTP- abstract redeneren (LTP-AR)

De LTP-AR is ontwikkeld als alternatief voor de test voor niet-verbale abstractie (TNVA). Deze tests meten het abstracte redeneervermogen bij kandidaten met een hoog opleidingsniveau (Drent, 1965, p. 502). In tegenstelling tot de LTP-AR bestaat de TNVA uit 45 items. De verdere opzet en duur van de test is hetzelfde. De LTP-abstract redeneren bestaat uit 40 items en duurt 20 minuten. Bij deze test krijgt men figuren aangeboden, bestaande uit vier opgavenplaatjes met een gemeenschappelijk kenmerk en zes antwoordplaatjes waarvan er twee hetzelfde kenmerk als de opgavenplaatjes bezitten. De kandidaat dient de juiste antwoordplaatjes te vinden.

LTP-verbaal redeneren (LTP-VR)

De LTP-verbaal redeneertest is als alternatief ontwikkeld voor een test die het verbaal-analytische vermogen van kandidaten met een hoog opleidingsniveau meet, de VAT-VA (Drent & Van Wieringen, 1969, p. 503). De VAT-VA is een van de drie subtests van de VAT '69 (verbale aanleg testserie), waarbij VA staat voor verbale analogieën. De opzet van de LTP-verbaal redeneren is gelijk aan de opzet van de VAT-VA. De test duurt 25 minuten en heeft 40 items. Er worden analogieopgaven aangeboden. Dit gebeurt in de vorm $A : B = C : D$, hierbij zijn de woorden op de plaatsen A en D weggelaten. Uit een serie alternatieven dient men het juiste alternatief te kiezen, zodat er een geldige structuur ontstaat.

Differentiële aanleg test '83 (DAT '83)

Deze test bestaat in zijn geheel uit negen subtests (Evers & Lucassen, 1991, p. 279). De Raio-kandidaten maken echter maar een subtest uit de DAT '83, namelijk de subtest zinnen. Hierbij meet men of de kandidaat (taal)vaardig genoeg is om goed te beoordelen of de opbouw van de gepresenteerde Nederlandse zinnen juist is. In totaal bestaat deze subtest uit 60 items, welke de kandidaat dient te beoordelen.

Syllogismentest

In deze test krijgt men telkens twee uitspraken voorgelegd. Uit deze twee uitspraken volgt een logisch verband. Het is aan de kandidaat om dit logische verband aan te geven. Er dient een logische conclusie getrokken te worden aan de hand van de uitspraken. Er worden vier alternatieven aangeboden waarvan er een juiste gekozen moet worden.

2.1.3 Voorgesprek

In deze fase van het selectieproces wordt men getoetst op enkele (persoonlijkheids-) basisvaardigheden. Er wordt hier gelet op de mondelinge uitdrukkingsvaardigheden, de wijze van optreden en evenwichtigheid. Bij dit gesprek zijn een lid van de Raio-selectiecommissie en een adviseur van het LTP aanwezig. Zij hanteren voor het gesprek een gestructureerde methode. Dit houdt in dat het gesprek volgens vaste regels verloopt. Op deze manier is het makkelijker om aantekeningen te maken en deze aantekeningen te vergelijken. Als de getoetste vaardigheden goed beoordeeld worden, kan men door naar de volgende fase. Bij een onvoldoende score wordt men afgewezen.

2.1.4 Assessment

Tijdens het assessment vindt een uitgebreid persoonlijkheidsonderzoek plaats. Dit wordt uitgevoerd door het LTP. Dit onderzoek vindt zowel mondeling als schriftelijk plaats. Als men een negatieve beoordeling heeft op het assessment kan er door de commissie alsnog besloten worden om de kandidaat uit te nodigen voor het selectiegesprek. Dit is daardoor het enige onderdeel waarbij men met een negatief resultaat wel door kan gaan naar de volgende fase van het proces.

Het assessment bestaat uit verschillende onderdelen: een interview met een adviseur van het LTP, het maken van enkele persoonlijkheidsvragenlijsten, een rollenspel en een simulatie. Deze onderdelen zullen hieronder verder besproken worden.

Interview

Het interview wat plaatsvindt tijdens het assessment, wordt gehouden aan de hand van het sollicitatieformulier, welke door de kandidaat is ingevuld. De psycholoog die aanwezig is bij het interview probeert aan de hand van de informatie uit het sollicitatieformulier én aan de hand van de gedragingen van de kandidaat een nauwkeurig beeld te krijgen van de persoon. Er wordt vooral gelet op hoe de kandidaat omgaat met stress en hoe hij omgaat met kritiek. Daarnaast wordt er gekeken naar de motivatie van de kandidaat, de evenwichtigheid, het doorzettingsvermogen, het samenwerken, het beeld van de Raio-opleiding en de kritische blik van de kandidaat. Hier wordt naar gekeken aan de hand van de gevolgde opleiding van de kandidaat en de werkervaring die men heeft opgedaan. Tijdens het gesprek maakt de psycholoog aantekeningen op een observatieformulier.

Persoonlijkheidstesten

Een ander onderdeel van het assessment is het maken van enkele persoonlijkheidstesten. In totaal dient de kandidaat vijf persoonlijkheidstesten te maken. De vijf testen welke momenteel gebruikt worden zijn:

- Schaal voor Interpersoonlijke Waarden (SIW); bij deze test wordt vooral gemeten wat mensen belangrijk vinden in het sociale verkeer met andere mensen (Drenth & Kranendonk, 1973, p. 361);
- LTP-vragenlijst (LTPV); deze vragenlijst is bedoeld om de gedragsstijl van de persoon redelijk te kunnen omschrijven;
- WIMAS; hier wordt gekeken naar manipulatief gedrag. Er wordt gekeken hoe de kandidaat, in relatie tot anderen, doelen tracht te bereiken (Altink-van den Berg & Akkerman, 1989, p. 563);
- LTP-werkstijlen (LTP-WS); bij deze test worden verschillende soorten motivatie gemeten;
- NEO-personality index (NEO-PI-R); dit is een veel gebruikte persoonlijkheidstest om de vijf domeinen ("Big Five") van de persoonlijkheid te meten. Deze vijf domeinen zijn: neuroticisme, extravertie, openheid, altruïsme en consciëntieusheid (Hoekstra, Ormel & Fuyt, 1996, p. 444).

Simulatie en rollenspel

Bij de simulatie en het rollenspel tijdens het assessment wordt men beoordeeld op verschillende criteria. Dit zijn de criteria denken, voelen en kracht. Ook wordt er gescoord op de effectiviteit van het gesprek.

Bij de simulatie krijgt de kandidaat de gelegenheid om in 45 minuten een dossier te bestuderen met achtergrondinformatie over een keuzeprobleem. Na het bestuderen krijgt de kandidaat vijf minuten de tijd om zijn keuze te presenteren en toe te lichten. Dit doet hij voor twee psychologen. Na deze vijf minuten moet hij nog kritische vragen beantwoorden over zijn keuze en de gemaakte keuze verantwoorden. Dit neemt ongeveer tien minuten in beslag. De aanwezige psychologen vullen hierna een beoordelingsformulier in. De kandidaat dient na afloop zijn advies voor het probleem op papier te zetten. Zo kan de kandidaat ook beoordeeld worden op zijn schriftelijke uitdrukkingsvaardigheden.

Bij het rollenspel heeft de kandidaat een kwartier om een situatie en het doel van het gesprek voor te bereiden. Hierna zijn er twintig minuten uitgetrokken voor een gesprek met een psycholoog of een acteur. Zij voeren vanuit een bepaalde rol het gesprek. Bij het gesprek is nog een andere psycholoog aanwezig die het gesprek observeert en aantekeningen maakt. Na het gesprek vullen de gespreksvoerder en de psycholoog beide een beoordelingsformulier in.

Beoordeling

De psycholoog die aanwezig was bij het interview, vat alle resultaten van het assessment samen. Hierin staat een advies over de kans van slagen die de kandidaat heeft met betrekking tot het afronden van de Raio-opleiding.

De Raio-commissie ontvangt deze beoordeling. Een negatieve score houdt echter niet in dat de kandidaat niet door kan gaan naar de volgende fase. De Raio-commissie kan zelf alsnog de beslissing maken om de kandidaat uit te nodigen voor het commissiegesprek. Deze beslissing kan gebaseerd worden op de resultaten die de kandidaat in de eerdere onderdelen van het proces behaald heeft. Kandidaten met een positieve score op het assessment gaan altijd door naar de volgende fase.

2.1.5 Commissiegesprek of selectiegesprek

Dit is de laatste fase van het selectieproces. In deze fase hebben de kandidaten een gesprek met drie leden van de Raio-selectiecommissie. Voordat het gesprek plaatsvindt moet men eerst schriftelijk enkele juridische vragen beantwoorden, waar in het gesprek verder op in wordt gegaan. In deze fase wordt de kandidaat getoetst op het algemeen juridisch inzicht, de motivatie, het functiebeeld van de kandidaat en de maatschappelijke belangstelling en betrokkenheid.

2.2 Raio-opleiding

Om een duidelijk beeld te krijgen van het geheel, zal kort de weg beschreven worden voordat men tot rechter benoemd kan worden. Er zijn twee wegen die naar de opleiding tot rechter leiden. Als men geen of minder dan zes jaar juridische werkervaring heeft kan men toegelaten worden tot de Raio (Rechterlijke Ambtenaar In Opleiding)-opleiding. Als men tenminste 6 jaar juridische werkervaring heeft stroomt men in als 'buitenstaander'. Hierbij komt de kandidaat terecht bij de CALRM (Commissie Aantrekken Leden Rechterlijke Macht) of bij de selectiecommissie voor het openbaar ministerie.

Omdat er in dit verslag alleen gekeken zal worden naar de Raio-selectieprocedure, zal alleen de Raio-opleiding kort besproken worden (zie voor een schematische weergave van de opleiding figuur 2.1). Echter voordat men toegelaten wordt tot deze opleiding dient men een streng selectieproces te doorlopen, welke 6 tot 8 weken duurt. Dit selectieproces is in de voorgaande paragraaf uitgebreid besproken.

Na het goed doorlopen van het selectieproces wordt men toegelaten tot de Raio-opleiding (Raad voor de Rechtspraak, 2003). Deze begint tweemaal per jaar, één keer in april en één keer in oktober. Gedurende het hele jaar zijn hiervoor selectierondes.

De eerste vier jaar van de opleiding wordt ook wel een binnenstage genoemd. Tijdens de eerste drie jaar van deze binnenstage is er een theoretisch deel waarbij men specifieke kennis verwerft. Daarnaast volgen de kandidaten een basiscurriculum bij een arrondissementsrechtbank en bij een parket.

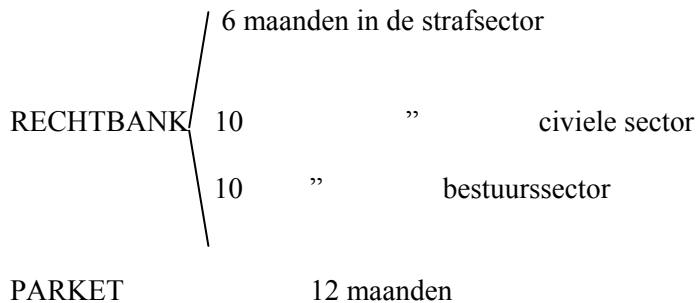
De eerste zes maanden fungeert men als secretaris in de strafsector en daarna fungeert men tien maanden als secretaris in de civiele sector. Als afsluiting van de rechtbankstage dient de kandidaat tien maanden als secretaris in de bestuurssector te werken. Na de rechtbankstage doet de kandidaat twaalf maanden dienst als plaatsvervangend officier van justitie bij een parket. Dit wordt de parketstage genoemd. Als men deze parketstage heeft afgerond, dient de kandidaat een keuze te maken voor de rechtbank of voor het parket.

Deze keus is bepalend voor de invulling van het vierde jaar, waarin een keuzestage plaatsvindt. Nadat de keuze gemaakt is, vindt een keuzestage plaats van tien maanden. Hierbij is men werkzaam als rechter (plaatsvervanger) bij een rechtbank of bij het parket als (plaatsvervangend) officier van justitie. Na de keuzestage is er een buitenstage. Deze stage duurt twee jaar en vindt plaats binnen de rechtspraktijk maar buiten de rechterlijke macht. Hier doet de kandidaat zelfstandig juridisch werk.

Als men de binnenstage en de buitenstage goed vervuld heeft, rond men de Raio-opleiding af. Na het goed afronden van de Raio-opleiding volgt een benoeming tot gerechtsauditeur of substituut officier bij een arrondissementsrechtbank of -parket. Na een jaar kan de kandidaat benoemd worden tot rechter of officier van justitie.

3 JAAR PRAKTIJKSTAGE

waarvan:



10 MAANDEN PRAKTIJKSTAGE (afhankelijk van keuze)

- RECHTBANK

KEUZESTAGE:

of

- PARKET van de OFFICIER VAN JUSTITIE

2 JAAR BUITENSTAGE

Figuur 2.1. Schematisch overzicht van de Raio-opleiding

3. Testgebruik bij allochtonen

Zowel in Nederland als in andere landen is er veel onderzoek gedaan naar het testen van etnische minderheden, in selectiesituaties, maar ook in andere situaties. Een veelbesproken test is de cognitief/analytische test ook wel CAT genoemd. De uitkomsten van de verschillende onderzoeken vertonen veel overeenkomsten in de resultaten, echter de verklaringen zijn zeer uiteenlopend. Om een duidelijk beeld te krijgen van de resultaten en de verklaringen is een literatuurstudie uitgevoerd. In de volgende paragrafen zullen enkele belangrijke onderzoeken en de bijbehorende resultaten besproken worden.

3.1 Algemeen

Enkele jaren geleden heeft er zich in Nederland een heftige discussie afgespeeld in 'De Psycholoog', vakblad van het Nederlands Instituut van Psychologen (NIP), over het testen van allochtonen. Dit begon met een artikel van Evers en Te Nijenhuis (1999) waarin zij een overzicht geven van de onderzoeken naar testbias in psychologische tests. De conclusie welke hier uit voortvloeit is dat testbias nauwelijks voorkomt en dat de resultaten van selectieprocedures hier dus niet substantieel door beïnvloed worden.

Door een werkgroep samengesteld door het Landelijk Bureau voor Racismebestrijding en het NIP is onderzocht in hoeverre tests discriminerend waren voor allochtonen. De uitkomst hiervan was dat de meeste tests die gebruikt worden in de selectiepraktijk mogelijk niet geschikt zijn voor het gebruik bij allochtone groepen. Dit omdat er in alle tests ethnocentrisme werd aangetroffen, wat inhoudt dat er impliciete of expliciete verwijzingen naar de Nederlandse cultuur voorkomen in de test. Ook werd door deze werkgroep benadrukt dat de tests eerst gevalideerd dienen te worden voor allochtonen, voordat tot het gebruik hiervan kan worden overgegaan.

Evers en Te Nijenhuis (1999) stellen dat, ondersteund door de consensus in de literatuur, tests goed, maar niet perfect, te gebruiken zijn voor assessment van allochtonen. De capaciteitentest bezit geen bias. Deze voorspelt slechts de prestaties van de persoon die de test doet. Het kan zijn dat de gemiddelde score van de ene groep lager ligt dan bij de andere groep en dit verschil in score niet veroorzaakt wordt door de test zelf. Bij tests met een taalcomponent kan hooguit een kleine testbias optreden. Hierdoor kan geconcludeerd worden dat bovengenoemde auteurs pleiten voor het gebruik van de standaard ontwikkelde tests bij allochtonen. Als zij lager scoren bezitten zij niet voldoende competenties om goed te functioneren in de functie waarvoor geworven wordt.

Van den Berg en Van Leest (1999) reageren op bovengenoemde door te pleiten voor een verantwoord testgebruik bij allochtonen. De beheersing van de Nederlandse taal speelt namelijk een zeer belangrijke rol. Hierdoor is het gebruik van een taaltest zeer aan te bevelen. Mocht er twijfel ontstaan dan kan men er ook voor kiezen om een test mondeling af te nemen, hierbij kan men nuttige informatie verkrijgen over antwoordstrategieën en basisvaardigheden. De beheersing van de Nederlandse taal blijkt een belangrijk punt te zijn, men moet namelijk wel de instructies en de vragen van de test zelf begrijpen.

Een ander belangrijk punt, wat genoemd wordt door Van den Berg en Van Leest (1999) is de bekendheid met psychologische tests. Een allochtoon kan hiermee minder bekend zijn dan autochtonen. Zo kan een allochtoon andere antwoordstrategieën hanteren waardoor hij/zij relatief lager scoort op de test.

Wil men psychologisch onderzoek doen bij allochtonen, dan moet dit zeer voorzichtig plaatsvinden. Men moet zowel letten op de testkeuze als op de interpretatie van de resultaten (Van den Berg en Van Leest, 1999).

De bovenstaande punten genoemd door Van den Berg en Van Leest trachten Te Nijenhuis en Evers (2000) te weerleggen. Zij stippen aan dat testbias nauwelijks voorkomt bij eerste en tweede generatie allochtonen, hierdoor is een herziening van de selectieprocedures niet nodig. Daarentegen stemmen zij wel in met het standpunt dat men voorzichtig moet zijn met het testen van allochtonen. Er dient meer bekendheid gegeven worden aan het fenomeen tests. Daarnaast blijkt dat allochtonen testinstructies niet altijd goed begrijpen. Dit is op te lossen door ze extra op gang te helpen of door het geven van extra oefeningen, bijvoorbeeld oefenopgaven. Door de COTAN is gesteld dat testtraining gegeven mag worden, mits dit de validiteit van de tests niet aantast.

Te Nijenhuis en Evers (2000) zijn tegen het aanpassen van de testbeoordeling, zodat men met een lagere score ook toegelaten kan worden, dit wordt het zogenaamde lager leggen van de meetlat genoemd. Deze procedure leidt tot lagere kwalificaties en lagere prestaties, waardoor men minder geschikte personen toch aanneemt voor een bepaalde functie.

Tellegen (2000) reageert op bovenstaande artikelen door het aandragen van enkele aspecten welke niet eerder besproken zijn. Zo geeft hij aan dat de Test-screeningscommissie voor een groot aantal tests tot de conclusie gekomen is dat de instructies en/of het testmateriaal aangepast dienen te worden op grond van etnocentrisme. Zoals eerder gezegd komt dit in veel tests voor.

Een ander belangrijk punt genoemd door Tellegen (2000), is dat wat gemeten wordt in een test in sterke mate afhankelijk is van taal, cultuur en schoolervaring. Het is dus verkeerd om tests voor allochtonen te baseren op de Nederlandse cultuur, taal en schoolervaring. De volgende vraag is hierom zeer belangrijk: Wanneer is het gebruik van traditionele tests verantwoord en voor welke testonderdelen is het gebruik verantwoord?

Ook geeft Tellegen (2000) aan dat er in de Algemene Standaard Testgebruik, opgesteld door het NIP (2000), vereist wordt dat er onderzoek wordt gedaan naar de taalvaardigheid en of deze invloed heeft op de testuitslag van niet-Nederlandstaligen. Mocht dit het geval zijn dan dient er een alternatief voor deze test gezocht te worden. Evers en Te Nijenhuis stelden echter, volledig in tegenstrijd met bovenstaande, dat de traditionele intelligentietests bij allochtonen gehanteerd kunnen worden en het gebruik van alternatieve tests volledig overbodig is.

3.2 Cognitief/analytische test

In de bovenstaand paragraaf zijn zaken besproken met betrekking tot het testen van allochtonen in het algemeen. Het gaat over alle ontwikkelde tests, die een allochtoon voorgelegd kan krijgen. Ook is er over het gebruik van de cognitief/analytische test bij verschillende bevolkingsgroepen veel gepubliceerd.

De cognitief/analytische test (CAT) wordt veel gebruikt in veel verschillende situaties, zowel bij selectieprocedures als bij andere situaties (Ellis en Ryan, 2003). Dit doet men omdat de CAT gezien wordt als een belangrijke predictor (voorspeller) voor vele verschillende criteria. Zo wordt er gezegd dat er een sterke link is met werkprestatie (Schmidt, 2002, Ployhart, Ziegert en McFarland, 2003, Outtz, 2002 en Schmidt en Hunter, 1998).

Ook wordt er gezegd dat scores op de cognitief/analytische test sterk samenhangen met de prestaties op school en alledaagse taken waarbij actieve informatieverwerking vereist is (Murphy, Cronin en Tam, 2003). De Raad voor de Rechtspraak heeft er om bovenstaande redenen voor gekozen om deze test als eerste fase van het selectieproces te gebruiken. Dit omdat het een goedkope manier van selecteren is en er redelijk veel mensen bij de cognitief/analytische test afvallen. Gemiddeld valt 50% van de groep kandidaten af bij de cognitief/analytische test (Cohen, 2003).

Het probleem bij de CAT blijkt echter te zijn dat er onbedoelde verschillen in de gemiddelde testcores zijn bij verschillende bevolkingsgroepen. Zo geven Te Nijenhuis, De Jong, Evers en Van der Flier (2004) aan dat de scores van allochtonen en autochtonen grote verschillen tonen, in het nadeel van de allochtonen. Andere auteurs geven eveneens aan dat er verschillen zijn in testcores bij verschillende groepen (Ellis en Ryan, 2003; Murphy et al., 2003 en Ployhart et al., 2003). Outtz (2002) geeft aan dat de scores gemiddeld ongeveer een standaarddeviatie verschillen. Voor deze verschillen in testcores worden enkele mogelijke oorzaken aangewezen (Ellis en Ryan, 2003 en Helms, 1992).

Sommigen wijten de verschillen aan een verschil in genetische basis, anderen zeggen dat de basis van de verschillen veroorzaakt wordt door de omgeving (economisch, sociaal en educatie) en een derde groep beweert dat de redenen van verschillen meer gerelateerd zijn aan de test en testsituatie. Reynolds (2000) geeft zelfs vier mogelijke verklaringen voor de verschillen in testcores, namelijk bovenstaande drie en de vierde is een combinatie van de effecten van de genetische factoren en omgevingsfactoren.

Als het verschil in testcores gerelateerd is aan de test en de testsituatie zijn er verschillende factoren welke de testcores kunnen beïnvloeden. Zo noemen Te Nijenhuis en Van der Flier (2004) vijf belangrijke culturele factoren welke een rol spelen op de uitkomsten van de test. Deze factoren zijn:

1. het begrijpen van de testinstructies;
2. bekendheid met het testmateriaal;
3. ervaring met de taak van de test;
4. ervaring met werken onder tijdsdruk;
5. testhouding, motivatie, concentratie, angst om te falen en sociale wenselijkheid.

Hoewel het onwaarschijnlijk lijkt dat deze factoren invloed hebben op de testresultaten van zulke hoogopgeleiden, zou dit toch wel het geval kunnen zijn (interview: Tillema, 2005). Deels is dit ook afhankelijk van wanneer de desbetreffende persoon naar Nederland is gekomen. Vooral de eerste drie factoren zijn zaken die men niet expliciet leert, maar waar men wel ervaring in opdoet bij de eindtoets van de basisschool en later op de middelbare school en het hoger onderwijs. Als men dit niet allemaal heeft meegemaakt, kan dit een gemis betekenen. De laatst genoemde factor kan sterk afhankelijk zijn van de cultuur, waardoor de gedachte van allochtonen bij het maken van de test heel anders is dan bij autochtonen.

Ellis en Ryan (2003) noemen nog enkele andere oorzaken, namelijk culturele bias en het effect van stereotypen op de prestatie van etnische minderheden. Ook worden de strategie welke gehanteerd wordt bij het maken van de test, de voorbereiding voor de test en test-taking self-efficacy genoemd als mogelijke zaken die effect hebben op de testcores.

Uit onderzoek is gebleken dat bovenstaande zaken de verschillen in testscores deels kunnen verklaren, maar geen van deze factoren kan of kunnen het totale verschil van scores volledig verklaren. Er wordt slechts een deel verklaart (Ployhart et al., 2003 en Ellis en Ryan, 2003).

Kort gezegd kan er geconcludeerd worden dat er veel onderzoek gedaan is naar de testverschillen tussen verschillende bevolkingsgroepen, maar deze verschillen zijn tot nu toe niet volledig verklaard door bovengenoemde punten. Er bestaat echter nog een mogelijkheid. Het kan zijn dat de verschillen in testscores niet aan de testnemer of testafnemer liggen, maar dat deze veroorzaakt worden door de test zelf. Het zou kunnen dat er fouten in de test zitten.

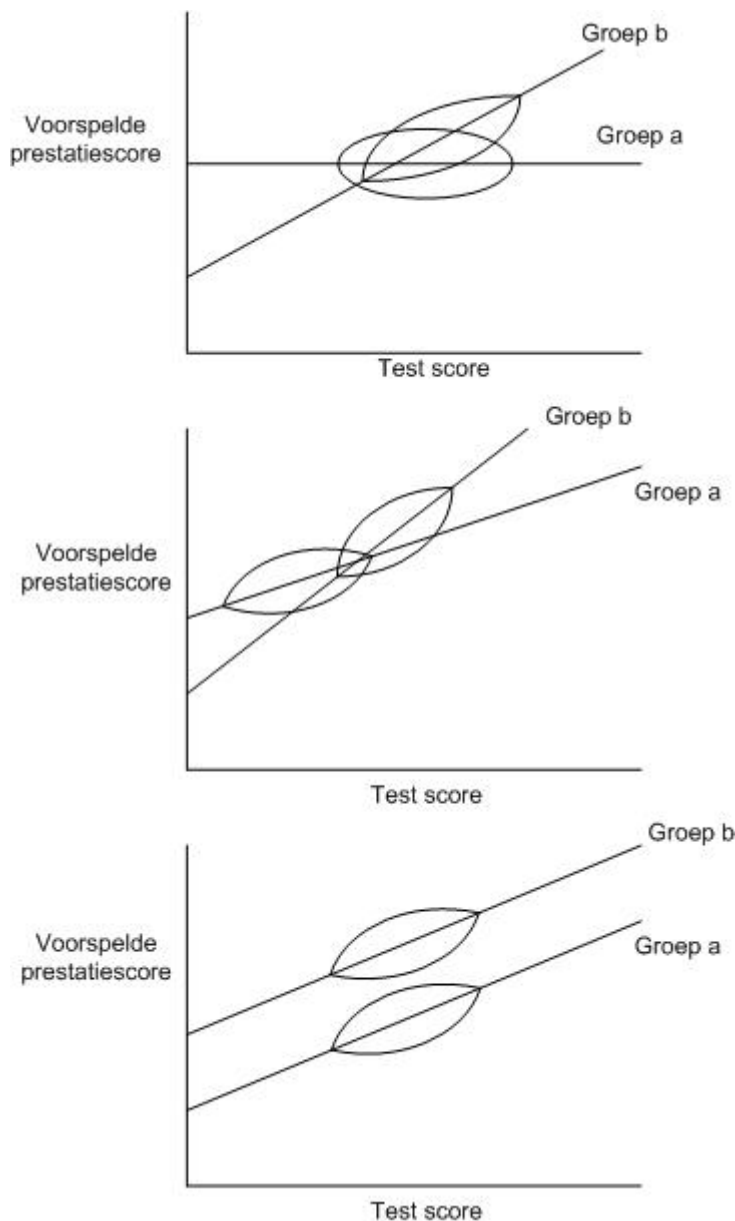
3.3 Bias

Hoewel het in dit onderzoek, bij een gebrek aan gegevens, niet mogelijk is om te kijken of de test gebiased is, zal dit echter wel kort besproken worden. Dit omdat het relevant is voor het testen van allochtonen.

Fouten in tests wordt ook wel testeerlijkheid (testfairness) of bias genoemd. Dit houdt in dat een test onbedoeld discrimineert tussen verschillende bevolkingsgroepen (Berry, 2003). Zo kunnen tests zich richten op bepaalde bevolkingsgroepen, waarbij men zich vooral richt op de meerderheidsgroepen (blank en middenklasse). Hierdoor scoren minderheidsgroepen lager. De tests kunnen op verschillende manieren oneerlijk zijn (Berry, 2003). De test kan valide zijn voor de meerderheidsgroep, maar niet voor de minderheidsgroep. Dit wordt single-group validity genoemd. Een andere mogelijkheid is differentiële validiteit, waarbij de scores van de ene groep iets anders betekenen dan de testscores van de andere groep.

Regressieanalyse kan gebruikt worden om te kijken of een test onbedoeld onderscheid maakt tussen verschillende groepen. Er kan gekeken worden naar de regressielijnen; de helling en het intercept. Dit is te zien in de figuren 3.1. a t/m c. In alle drie figuren is op drie verschillende manieren duidelijk te zien dat eenzelfde testscore voor de twee groepen een verschillende voorspellende prestatiescore hebben. Dit ten nadele van een van de twee groepen.

Humpreys (1986) noemt dezelfde manieren om bias te bespeuren. Hierbij maakt hij onderscheid tussen twee verschillende soorten bias, itembias en testbias. Deze twee vormen van bias kunnen volkomen afhankelijk van elkaar voorkomen. Bij itembias wordt gekeken of er bepaalde items in de test zijn welke zorgen voor het onderscheid tussen verschillende groepen. Bij testbias wordt er gekeken of er in de totale score op de test onderscheid gemaakt wordt tussen verschillende groepen. Hierbij geeft Humpreys (1986) echter aan dat bias een erg emotioneel geladen woord is, dit kan door sommigen verkeerd opgevat worden. Het is beter om de term 'onderscheid' te hanteren, omdat deze term eveneens duidelijk maakt wat bedoeld wordt.



Figuur 3.1. Verschillende voorbeelden van bias (Berry, 2003).

3.4 Kritische kanttekeningen

Uit de besproken onderzoeken in de voorgaande paragrafen blijkt dat de vele verklaringen die er zijn voor de verschillen in test scores deze verschillen niet volledig dekken. Onderzoekers zeggen dat er geen fouten in de test zitten, maar dat het ook niet mogelijk is om de verschillen in test scores volledig te verklaren met behulp van de andere aangedragen verklaringen.

Ployhart et al. (2003) hebben een model opgesteld met factoren welke een rol kunnen spelen in de beïnvloeding van de test scores voor de cognitief/analytische test. Ze hebben alle mogelijke factoren, die waren opgenomen in het model, onderzocht en de bijdrage van deze factoren opgenomen in de herziene versie van het model.

Uit dit model komt naar voren dat de verschillen in testcores niet volledig door deze factoren verklaard worden. In het huidige onderzoek is het echter niet mogelijk om de factoren welke in dit model opgenomen zijn te onderzoeken. Daarom zal dit genoemde model niet verder besproken worden.

Schmidt (2002) geeft aan dat het doel wat sommige bedrijven momenteel nastreven met betrekking tot het sociale beleid en een gelijke representatie van alle bevolkingsgroepen in alle banen onhaalbaar is. Zelfs met tests die geen onbedoeld onderscheid maken en eerlijke selectiemethoden is dit een onrealistisch doel. De verschillen in testcores bewijzen dat er verschillen zijn tussen verschillende groepen in baanrelevante ervaringen en vaardigheden. Een representatieve verdeling van de bevolking in het bedrijf zou hierdoor leiden tot lagere kwalificaties en prestaties van het bedrijf (Te Nijenhuis en Evers, 2000). Sackett (2001) geeft hierbij aan dat het onmogelijk is om te verwachten dat zowel de prestatie van het bedrijf als het aantal allochtonen wat werkzaam is bij dat bedrijf te maximaliseren.

Representativiteit in een organisatie is belangrijk, zeker voor de Rechterlijke Macht, maar het is niet het belangrijkste (Sloot, 2004). Elke rechter zou juist oordelen moeten kunnen vormen. Representativiteit van de Rechterlijke Macht is echter wel wenselijk. Dit omdat de beslissingen, genomen door een diverse groep, waarin verschillende bevolkingsgroepen gerepresenteerd zijn, vaker geloofwaardiger en kwalitatief beter zijn. Wil men dit bereiken is het echter niet de bedoeling dat men dit op een onjuiste manier gaat doen. Een 'extreme' voorkeursbehandeling is bijvoorbeeld uit den boze. Een voorkeursbehandeling zou alleen gerechtvaardigd zijn bij een gelijke, voldoende testscore, waarbij de 'minderheid' doorgaat en de andere persoon niet.

3.5 Conclusies

Bovenstaand is te lezen dat er zeer uiteenlopende ideeën zijn betreffende het testen van allochtonen. Kort samengevat kan er het volgende gezegd worden; men dient zeker voorzichtig te zijn met het afnemen van tests bij allochtonen. Het is erg belangrijk dat men kijkt naar de taalvaardigheid. Als deze onvoldoende blijkt te zijn, is de kans groot dat de testresultaten een vertekend beeld vormen, waardoor men geen correcte conclusies kan trekken uit de resultaten. Men kan kijken of extra tests of trainingen gewenst of noodzakelijk zijn, voordat men overgaat tot het afnemen van de traditionele test bij een allochtoon.

Daarnaast zou men er voor kunnen kiezen om de traditionele tests te valideren en aan te passen zodat deze te gebruiken zijn bij allochtonen. Een andere mogelijkheid is om tests te gebruiken welke speciaal ontwikkeld zijn voor allochtonen. Naast deze punten is het zeker aan te raden om het etnocentrisme uit de traditionele tests te halen, zodat deze tests bij alle bevolkingsgroepen te gebruiken zijn.

4. Methoden

In dit deel zal beschreven worden hoe de eerder gestelde onderzoeksvraag: *Zijn er onbedoelde discriminatoire effecten in de cognitief/analytische test tussen alloctonen en autoctonen?* beantwoord zal worden en hoe de bijbehorende hypothesen getoetst zullen worden. De opgestelde hypothesen zijn:

1. *Er wordt verwacht dat alloctonen die de uitgebreide versie van de cognitief/analytische test maken een grotere kans van slagen hebben voor deze fase van het selectieproces;*
2. *Er wordt verwacht dat er een verschil is in de scores tussen de verschillende subtests;*
 - a. *Er wordt verwacht dat de score op de DAT '83 (zinnen) significant lager ligt dan het gemiddelde op de andere testscores;*
 - b. *Ondanks dit wordt er verwacht dat de score op de LTP-verbaal redeneren test lager zal liggen dan de score op de andere redeneer test (LTP-abstract redeneren);*
3. *Er wordt verwacht dat er een vooruitgang te zien is in de testscores over de jaren heen.*

Het hoofdstuk is in twee delen op te splitsen. Eerst zal beschreven worden hoe de data verkregen worden en welke testdata gebruikt zullen worden voor het beantwoorden van de vraag. Als dit gedaan is, zal de gehanteerde procedure beschreven worden.

4.1 Testgegevens

De data zijn op twee verschillende manieren verkregen. Enerzijds heeft dit op een kwantitatieve manier plaatsgevonden, anderzijds is dit kwalitatief gebeurd.

De kwantitatieve data zijn verkregen door enkele bezoeken aan het LTP te Amsterdam. Tijdens deze bezoeken heeft er eenmaal een gesprek met een Marokkaanse kandidaat plaatsgevonden, voordat deze persoon de cognitief/analytische test voorgelegd zou krijgen. Bij dit gesprek werd geobserveerd en is er nauwelijks aan het gesprek deelgenomen. Hiernaast is er ook een gesprek geweest met Dhr. Tillema over het selectieproces en de problematiek die het testen van alloctonen met zich mee brengt. Ook is er tijdens deze bezoeken duidelijk gemaakt hoe het selectieproces verloopt voor de Raio-kandidaten en hoe de tests worden afgenomen.

De kwalitatieve data zijn op verzoek van de Raad voor de Rechtspraak door het LTP opgezocht en aangereikt. Het gaat hier om de testscores van de cognitief/analytische test van alle alloctonen die zich hebben aangemeld voor de Raio-opleiding vanaf ongeveer april 2000 tot april 2005. De scores van de vier subtests en de uiteindelijke somscores zijn gegeven.

Voor het beantwoorden van de onderzoeksvraag heeft er een data-analyse van deze testgegevens plaatsgevonden. Een analyse van de testgegevens van autochtone kandidaten was niet nodig, daar dit reeds gedaan is door het LTP te Amsterdam. De gemiddelde test- en subtestscores zijn 500 ($z = 0$) met een standaarddeviatie van 100 ($\sigma = 1$).

De scores van alle allochtone kandidaten vanaf 2000 zijn gebruikt voor de analyse van de testgegevens. Dit zijn zowel de testgegevens van allochtonen die deelgenomen hebben aan de standaard cognitieve/analytische test als de testgegevens van allochtone kandidaten die de uitgebreide versie van de test hebben gedaan.

Voor het beantwoorden van de onderzoeksvraag is, zoals uit bovenstaande beschrijving op te maken is, gebruik gemaakt van de gehele populatie kandidaten voor de Raio-opleiding. Deze groep is echter opgesplitst naar een populatie autochtone kandidaten en een populatie allochtone kandidaten. In beide groepen werden alle gegevens gebruikt. Tevens is er, voor het kunnen beantwoorden van een van de gestelde hypothesen, een opsplitsing gemaakt binnen de allochtone groep. De groep is hiervoor opgesplitst naar een groep welke heeft deelgenomen aan de gewone cognitief/analytische test in Den Haag en een groep welke ervoor heeft gekozen om de uitgebreide versie van de test te doen bij het LTP te Amsterdam. Deze opsplitsing heeft echter alleen plaatsgevonden voor het beantwoorden van hypothese één: *"Er wordt verwacht dat allochtonen die de uitgebreide versie van de cognitief/analytische test maken een grotere kans van slagen hebben voor deze test"*.

Het is van belang om te weten dat kandidaten met een somscore van 500 of hoger doorgaan naar de volgende fase. Bij de autochtonen gaat er 50% door naar de volgende fase. Of dit ook het geval is voor allochtone kandidaten zal onderzocht worden in het volgende hoofdstuk.

4.2 Procedure

Zoals gezegd zal de onderzoeksvraag beantwoord worden door middel van data-analyse. Aan de hand van deze data-analyse is gekeken of er onderscheid gemaakt wordt tussen allochtonen en autochtonen. Er heeft een analyse van de data plaatsgevonden per subtest (verbaal redeneren, abstract redeneren, syllogismen en zinnen) en van de overallscore, ook wel de somscore genoemd. Deze somscore bestaat uit een gewogen gemiddelde van de scores op de subtests. Kort gezegd zal er begonnen worden met beschrijvende statistiek, er zal een beschrijving worden gegeven van de testdata.

Naast deze algemene beschrijvende statistieken zijn de data verder uitgewerkt. Er is gekeken of er een trend is te onderscheiden in de testscores. Dit houdt in dat er vergelijkingen plaats hebben gevonden tussen de testscores van de verschillende jaren. Dit werd gedaan per subtest, maar ook voor de somscore. Door hier naar te kijken kan er mogelijk iets gezegd worden over de ontwikkeling en de trend van de testscores van de allochtone kandidaten.

Ook heeft er een vergelijking plaatsgevonden van de scores per subtest. Er is gekeken of er op een van de subtests minder goed gescoord wordt dan op de andere tests. Verder zijn de kwantitatief verkregen data gebruikt voor het onderbouwen, verantwoorden en verklaren van de resultaten, maar zijn ze ook gebruikt voor het trekken van conclusies. Daarnaast ondersteunen ze, waar mogelijk, de gedane aanbevelingen.

Om goede uitspraken te kunnen doen over een mogelijk onbedoeld onderscheid in de cognitief/analytische test is het aan te raden om te kijken of de tests gebiased zijn. Helaas is dit niet mogelijk. Dit omdat alleen de testscores verkregen zijn, de zogenaamde predictorscores. Helaas zijn hier geen criteriumscores bij verkregen waardoor het niet mogelijk is om het geheel te koppelen en regressielijnen op te stellen. Deze zijn noodzakelijk om vast te stellen of er sprake is van testbias. Itembias kan eveneens niet vastgesteld worden in dit onderzoek omdat er geen itemscores verkregen zijn.

4.3 Verantwoording aanpak

Er is om verschillende redenen gekozen voor bovenstaande aanpak. Allereerst is er voor het analyseren van de data gekozen om zo een duidelijk beeld van de data te kunnen vormen. Daarnaast kunnen er, met behulp van de gegevens die hieruit naar voren komen, relevante en mogelijk statistisch significante conclusies getrokken worden. Dit is van belang voor het beantwoorden van de gestelde onderzoeksvraag en voor het toetsen van de hypothesen.

Er zal, zoals reeds eerder vermeld, niet ingegaan worden op het al dan niet gebiased zijn van de test. Dit omdat dit niet mogelijk is. Om hier uitspraken over te doen zijn er andere gegevens nodig waarover geen beschikking is in dit onderzoek.

5. Resultaten

In de onderstaande paragrafen zullen de resultaten besproken worden van de data-analyse. Eerst zal er in gegaan worden op enkele algemene gegevens van de data, de zogenaamde beschrijvende statistieken. Hierna zal ingegaan worden op de specifiekere data-analyse om de onderzoeksvragen te beantwoorden.

5.1 Beschrijving data

In totaal zijn er 93 allochtonen ($n = 93$) geweest die de cognitief/analytische test van 2000 tot nu hebben gemaakt. Van deze 93 mensen zijn er negen mensen geweest die ervoor gekozen hebben om de uitgebreide versie van de cognitief/analytische test te maken ($n = 9$). Helaas zijn er van deze mensen geen scores te achterhalen (missing values of missende waarden). Wel is het mogelijk geweest om te achterhalen hoeveel van deze mensen door zijn gegaan. Van de kandidaten die de uitgebreide test hebben gedaan zijn er twee doorgedaan naar de volgende fase van het selectieproces. Van de gehele populatie zijn er 20 kandidaten (inclusief de twee kandidaten van de uitgebreide test) door gegaan naar de volgende selectiefase.

Doordat er geen testcores bekend zijn van de kandidaten die de uitgebreide versie van de test hebben gedaan, kan er niet direct gekeken worden naar mogelijke verschillen in scores tussen de twee groepen. Hier moet echter bij vermeld worden dat, mochten de gegevens wel te achterhalen zijn geweest, een vergelijking tussen de verschillende tests erg moeilijk zou zijn geweest. Dit omdat de uitgebreide test uit heel andere tests bestaat en daardoor anders gescoord wordt. Door deze andere score en dus een andere betekenis van de scores, kan dit niet goed vergeleken worden. Er zal verderop alleen gekeken worden of men in een van de twee groepen een grotere kans heeft om door te gaan naar de volgende fase en wordt er niet gekeken naar verschillen in scores.

Voordat begonnen zal worden met de data-analyse ter beantwoording van de vragen, zullen eerst enkele algemene gegevens van de testcores gegeven worden. In tabel 5.1a zijn de algemene beschrijvende data te vinden, in tabel 5.1b zijn de beschrijvende statistieken van de subtests en de somscore per jaar terug te vinden. In deze tabellen zijn de gemiddelde scores van de subtests en van de gehele test te zien. Ook zijn de minimum en maximumscores en de standaarddeviaties terug te vinden. De scores zijn van de Raio-kandidaten vanaf 2000 tot 2005.

Tabel 5.1a. Overzicht beschrijvende statistieken Raio-scores.

	N	Minimum	Maximum	Gemiddelde	Std. Deviatie
Syllogistentest	84	200	683	452.48	100.89
Zinnen (DAT'83)	84	200	662	420.83	124.49
Verbaal redeneren	84	218	623	399.95	89.46
Abstract redeneren	84	200	648	442.36	105.73
Somscore	84	200	624	400.87	106.24
Valid N (listwise)	84				

Tabel 5.1b. Beschrijvende statistieken van de tests per jaar.

Jaar	Test	N	Minimum	Maximum	Gemiddelde	Std. deviatie
2000	Syllogismen	8	234	511	403.63	83.94
	Zinnen	8	200	518	365.75	117.69
	Verbaal redeneren	8	281	481	376.00	64.81
	Abstract redeneren	8	203	544	416.25	114.56
	Somscore	8	200	454	345.00	102.66
2001	Syllogismen	14	311	619	450.71	95.77
	Zinnen	14	200	552	406.43	109.14
	Verbaal redeneren	14	218	551	392.36	102.94
	Abstract redeneren	14	314	648	435.21	108.96
	Somscore	14	256	522	382.93	90.70
2002	Syllogismen	16	320	592	467.38	82.55
	Zinnen	16	224	604	469.44	112.02
	Verbaal redeneren	16	292	514	397.75	65.10
	Abstract redeneren	16	302	615	449.56	87.59
	Somscore	16	303	566	425.38	77.58
2003	Syllogismen	12	292	683	468.83	101.14
	Zinnen	12	214	591	427.83	138.86
	Verbaal redeneren	12	275	602	446.92	117.44
	Abstract redeneren	12	200	639	471.50	129.43
	Somscore	12	219	624	435.08	142.18
2004	Syllogismen	26	200	599	449.54	108.58
	Zinnen	26	200	662	388.38	133.85
	Verbaal redeneren	26	234	623	384.04	89.22
	Abstract redeneren	26	200	639	419.88	110.79
	Somscore	26	200	609	379.46	112.03
2005	Syllogismen	8	236	627	459.63	141.54
	Zinnen	8	340	609	498.88	79.89
	Verbaal redeneren	8	336	561	422.88	75.66
	Abstract redeneren	8	392	549	495.88	56.18
	Somscore	8	364	584	457.38	74.75

5.2 Analyse

In onderstaande paragraaf zal een uitgebreide analyse van de gegevens plaatsvinden. In tabel 5.2 is het overzicht gegeven van de totale groep en het aantal wat doorgaat naar de volgende fase van het selectieproces. Daarnaast is de groep opgesplitst naar de mensen die de standaard versie van de cognitief/analytische test gedaan hebben en de groep welke de uitgebreide versie van de cognitief/analytische test gedaan hebben. Hier staan eveneens de aantallen bij van de kandidaten welke door gaan naar de volgende fase. Omdat er geen testcores bekend zijn van de kandidaten welke de uitgebreide test gedaan hebben kan er geen uitgebreidere analyse en vergelijking plaatsvinden ter ondersteuning van hypothese één.

Tabel 5.2. Overzicht van het aantal kandidaten die doorgaan naar de volgende fase.

	N	Door naar volgende	
		fase	Percentage door
Totale groep	93	20	21.5 %
Standaard versie	84	18	21.4 %
Uitgebreide versie	9	2	22.2 %

Uit bovenstaande resultaten blijkt dat het verschil tussen de twee groepen en het percentage wat doorgaat naar de volgende fase minimaal is. Het verschil bedraagt slechts 0,8 %. Gezien het kleine aantal kandidaten welke de uitgebreide versie van de test gedaan hebben en het kleine aantal wat is doorgegaan naar de volgende fase, kunnen er geen uitspraken op significant niveau gedaan worden.

Voor het kunnen beantwoorden van de tweede hypothese is een vergelijkende analyse van de subtests noodzakelijk. Dit zal gedaan worden door een analyse van de variantie tussen verschillende groepen (ANOVA). De beschrijvende statistieken van de subtests zijn terug te vinden in tabel 5.1. Voor het beantwoorden van de tweede subhypothese van hypothese twee zal er naast de ANOVA gebruik gemaakt worden van de gepaarde T-toets. Dit wordt gedaan voor de vergelijking van de scores van de verbale redeneertest en de abstracte redeneertest.

ANOVA wijst uit dat er net geen significant verschil is tussen de DAT '83 (zinnen) en de overige tests ($F = 1.931$ en $p = .099$). Daarnaast zijn er tussen de overige subtests eveneens geen significante verschillen te zien. Wel blijkt uit de gepaarde T-toets, die uitgevoerd is voor de vergelijking tussen de verbale redeneertest en de abstracte redeneertest dat er een significant verschil is tussen deze twee tests ($T = -4.064$ en $p = .000$).

Geconcludeerd kan worden dat op grond van de gestelde hypothesen en de daarbij behorende resultaten dat het verschil tussen de abstracte redeneertest en de verbale redeneertest significant is. Daarnaast is er geen significant verschil te zien tussen de DAT '83 en de overige testscores. Tussen de overige tests zijn eveneens geen significante verschillen gevonden.

Voor het toetsen van de derde gestelde hypothese: *Er wordt verwacht dat er een vooruitgang te zien is in de testscores over de jaren heen*, wordt gebruik gemaakt van regressieanalyse. Er wordt een regressieanalyse uitgevoerd per subtest en voor de gehele test, de somscore.

Uit de regressieanalyse blijkt dat er geen significante vooruitgang is te zien. Dit geldt zowel voor de verschillende subtests als voor de somscore. Er is wel een kleine vooruitgang te zien, vooral in de totaalscore, maar deze is zeker niet significant. Het kan echter het geval zijn dat er wel degelijk een vooruitgang in de score is van allochtonen, maar dat hier afbreuk aan wordt gedaan door de lage scores uit 2004 (gemiddelde somscore is dit jaar 379.46). Hier kan echter niet met zekerheid een antwoord op gegeven worden.

5.3 Samenvatting resultaten

Over de resultaten kan gezegd worden dat het aantal allochtonen wat doorgaat naar de volgende fase aanzienlijk kleiner is dan het aantal autochtonen wat doorgaat naar de volgende fase (21,5% van de allochtonen tegen 50% van de autochtonen). Ook de allochtonen die ervoor kiezen om de uitgebreide versie van de cognitief/analytische test te maken, hebben geen significant grotere kans om door te gaan naar de volgende fase. Het verschil is slechts 0.8%. Aan de hand van dit verschil in percentage wat doorgaat naar de volgende fase, zou gezegd kunnen worden dat er in de test onbedoeld onderscheid gemaakt wordt tussen autochtonen en allochtonen.

Er zijn duidelijke verschillen in scores tussen de subtests te zien, zoals in de beschrijvende statistieken terug te vinden is. Het enige waargenomen significante verschil in dit onderzoek is tussen de verbale redeneertest en de abstracte redeneertest. De overige verschillen tussen de tests zijn niet significant.

Gezien de resultaten van de regressieanalyse kan er gezegd worden dat er geen significante ontwikkeling van de testscores te zien is. Er kan niet gezegd worden dat de kandidaten zich in de loop van de afgelopen vijf jaar zo hebben ontwikkeld waardoor de scores significant zijn toegenomen.

Op de onderzoeksvraag: "*Zijn er onbedoelde discriminatoire effecten in de cognitief/ analytische test tussen allochtonen en autochtonen?*" kan niet met zekerheid een juist antwoord gegeven worden. Uit de resultaten blijkt duidelijk dat er minder allochtonen doorgaan dan autochtonen. Dit zou kunnen door onbedoelde effecten in de test. Het is duidelijk te zien dat er in de scores op de subtest grote verschillen zijn, terwijl bij autochtonen de gemiddelde somscore, maar ook de gemiddelden van de subtests 500 zijn. Het kan zijn dat er minder allochtone kandidaten doorgaan doordat er onbedoelde discriminatoire effecten zijn, maar dit zou ook ergens anders door kunnen. Bijvoorbeeld doordat het verbale vermogen van allochtonen niet alleen in de Nederlandse taal, maar ook in hun moedertaal anders ontwikkeld is dan het verbale vermogen van autochtonen. Dit hoeft niet per definitie minder goed te zijn, maar anders, waardoor ze lager scoren op de verbale redeneertest en andere tests, omdat het op een andere manier getest wordt, waardoor voor de allochtone kandidaten op een verkeerde manier aanspraak wordt gedaan op het verbale vermogen. Ook kan het zijn dat de achterliggende gedachte of motivatie van de kandidaten bij het maken van de test anders is (Ryan, 2001), waardoor de testresultaten beïnvloed worden. Dit heeft echter niets met de test zelf te maken. Er kan natuurlijk ook sprake zijn van een combinatie van beide, dus dat er onbedoeld onderscheid wordt gemaakt door de test zelf, maar ook dat het verschil door de allochtone kandidaten zelf veroorzaakt wordt.

6. Discussie

In onderstaand hoofdstuk zullen de resultaten en conclusies gekoppeld worden aan reeds gepubliceerde onderzoeken. In het hierop volgende hoofdstuk zullen hierop aansluitend enkele aanbevelingen voor mogelijke vervolgonderzoeken gedaan worden.

6.1 Resultaten

De resultaten uit dit onderzoek brengen geen verrassende nieuwe zaken aan het licht. Veel van de resultaten zijn al eerder behandeld in eerder gepubliceerde artikelen, zowel in binnen- als buitenland. In dit onderzoek komt naar voren dat de gemiddelde score van de etnische minderheden (= 400,87) een standaarddeviatie lager ligt dan de gemiddelde score van autochtonen (= 500) op de test. Dit is in eerder onderzoeken al aan gegeven (Outz, 2002). Er zijn ook velen die verschillen in testcores tussen verschillende bevolkingsgroepen impliceren, maar hier geen getallen aan koppelen (onder andere Te Nijenhuis, De Jong, Evers en Van der Flier, 2004; Te Nijenhuis en Van der Flier, 2004; Helms-Lorenz, Van de Vijver en Poortinga, 2002).

In deze onderzoeken worden ook veel mogelijke oorzaken voor de verschillen in testcores aangegeven (Ryan, 2001; Ellis en Ryan, 2003; Murphy et al., 2003; Ployhart et al., 2003; Helms, 1992; Reynolds, 2000), zie hiervoor ook het hoofdstuk waarin de literatuur uitvoerig beschreven is (zie hoofdstuk 3). Gezien de verkregen data in dit onderzoek, was het echter niet mogelijk om uitvoerig onderzoek te doen in hoeverre deze factoren, genoemd in verschillende artikelen, invloed hebben op de scores van de allochtonen die de cognitief/analytische test gemaakt hebben.

De verschillen in scores die er tussen de verschillende subtests zijn, wordt niet direct verklaard door de literatuur. Wel kan het zijn dat de subtests aanspraak maken op verschillende aspecten van intelligentie, bijvoorbeeld op de verschillende niveaus van het drie-stratummodel van Carroll (Evers en Te Nijenhuis, 1999).

Dit houdt in dat men verschillende lagen van intelligentie (eerst algemene intelligentie en daarna meer specifieke intelligentie) kan aanspreken. Ook zou het kunnen zijn dat men bij de subtests aanspraak maakt op cognitieve of culturele factoren (Helms-Lorenz et al., 2002), wat zorgt voor de verschillen. Omdat er in de literatuur verder weinig is gezegd over verschillen in subtests, kunnen er geen uitspraken over deze verschillen gedaan worden.

De bevinding dat er geen significante ontwikkeling van de testcores te zien is, wordt niet direct ondersteunt door het onderzoek uitgevoerd door Te Nijenhuis et al. (2004). De resultaten van dit onderzoek tonen aan dat de verschillen tussen etnische minderheden en autochtonen op verschillende gebieden minder worden. Dit is echter bij de cognitief/analytische test in het huidige onderzoek niet het geval. Een mogelijke verklaring voor dit verschil kan zijn dat er door Te Nijenhuis et al. (2004) onderzoek is gedaan naar andere aspecten van kennis dan in dit onderzoek. Ook kan het zijn dat men in het onderzoek gebruik heeft gemaakt van andere testgegevens of een andere samenstelling van de geteste groep (bijvoorbeeld minder hoogopgeleiden). Momenteel is er geen significante vooruitgang in de testcores te zien, maar de kans bestaat dat dit in de komende jaren gaat veranderen.

6.2 Selectieproces

De mogelijkheid die de Raad voor de Rechtspraak aan allochtone kandidaten biedt, om een uitgebreide versie van de cognitief/analytische test te doen, is erg goed. Zo kan in ieder geval een van de genoemde factoren welke mogelijk invloed heeft op de testscore, namelijk het begrijpen van de testinstructie (Te Nijenhuis en Van der Flier, 2004), beperkt worden. In de uitgebreide versie van de test hebben de allochtonen meer tijd voor het doorlezen van de testinstructie, dit omdat deze test vaak op individueel niveau wordt gegeven, waardoor ze rustig de tijd kunnen nemen voor de instructie. Dit wordt door velen als positief ervaren (interview: Tillema, 2005). Bij de standaardtest is er geen tijd voor het uitgebreid doorlezen van de instructies, omdat het geheel in groepsverband wordt afgenomen en aan tijd gebonden is. Toch blijkt uit de resultaten dat deze extra tijd voor het doorlezen van de testinstructies geen grote invloed op de testscores heeft.

Een aanbeveling welke gedaan wordt door Van den Berg en Van Leest (1999) is om een taalttest vooraf te gebruiken bij belangrijke tests. Voor de allochtonen welke voor de uitgebreide versie van de cognitief/analytische test kiezen, is dit reeds het geval. Voordat zij met de cognitief/analytische tests beginnen, krijgen zij eerst de WLA (Woordenlijst allochtonen) voorgelegd door het LTP. Dit wordt gedaan ter inventarisatie van de taalkennis van de betreffende kandidaat.

Samenvattend kan gezegd worden dat er in de literatuur veel ondersteuning is voor de gevonden resultaten. Het is duidelijk dat er verschillen in testscores zijn. Om deze reden is het goed dat de mogelijkheid wordt aangeboden om een uitgebreide versie van de test te maken. Ook het voorleggen van de WLA aan de allochtone kandidaten die de uitgebreide versie van de cognitief/analytische test maken, kan veel verduidelijken betreffende de taalkennis van de kandidaat. Hiermee kan mogelijk een verband gelegd worden tussen de score op de WLA en de scores op de overige tests.

7. Aanbevelingen

In dit hoofdstuk zullen verscheidene punten aan bod komen. Eerst zullen enkele verbeterpunten genoemd worden van het huidige onderzoek. Aan de hand hiervan zullen ook enkele aanbevelingen gedaan worden voor mogelijke vervolgonderzoeken en voor andere vervolgstappen die de Raad voor de Rechtspraak zou kunnen nemen.

7.1 Tekortkomingen en aanbevelingen

Zoals gezegd kan er niet met zekerheid een antwoord gegeven worden op de vraag of er onbedoelde discriminatoire effecten aanwezig zijn in de cognitief/analytische test. Gezien de resultaten is de kans echter groot dat deze effecten aanwezig zijn. Om met meer zekerheid een antwoord te kunnen geven en om sterkere conclusies te kunnen trekken is het aan te raden om een grondigere analyse van de data uit te voeren. Het was in dit onderzoek niet mogelijk om met zekerheid vast te stellen of er daadwerkelijk een onbedoeld onderscheid werd gemaakt in de cognitief/analytische test tussen allochtonen en autochtonen of dat het gevonden verschil in testcores een andere oorzaak heeft. Om dit te kunnen onderzoeken is het noodzakelijk dat er meer data beschikbaar zijn. Dit houdt in dat het niet alleen wenselijk is om de beschikking te hebben over de subtestcores, zoals dit nu het geval was, maar ook de itemscores zouden hierbij van pas komen. Als er naar itemscores gekeken wordt kan men onderzoek doen naar mogelijke itembias. Door te kijken of itembias voorkomt in de gebruikte test kunnen de meer directe oorzaken van onderscheidend vermogen aangewezen worden. Als deze gevonden worden kan men deze oorzaken verhelpen, waardoor het onderscheidend vermogen tussen verschillende bevolkingsgroepen vermindert of zelfs helemaal verholpen wordt.

In dit onderzoek zou het mooi zijn geweest als de data van de kandidaten welke de uitgebreide versie van de test hebben gedaan, meegenomen waren in de analyse. Zoals eerder gezegd zou dit erg lastig zijn door de verschillen in test. Door het meenemen van deze data, zou men kunnen vergelijken of de scores op de uitgebreide versie minder ver onder de minimale grens van doorgaan liggen dan bij de allochtonen welke de standaardversie van de test maken het geval is. Ook hier zou het mooi zijn als itemscores verkregen waren voor een itemanalyse.

Kort samengevat kan geconcludeerd worden dat het voor een volgend onderzoek wenselijk is om de beschikking te hebben over meer gegevens van de kandidaten, zowel van de autochtone als de allochtone, om op die manier de data uitgebreider te kunnen analyseren en sterkere conclusies te kunnen trekken. Belangrijk is dat men zich realiseert dat dit een groot aantal gegevens op zal leveren, als deze data verkregen zijn. Daarnaast heeft het LTP aangegeven dat het erg lastig is om aan itemscores van de tests te komen in verband met privacy en vernietiging van de dossiers na 3 jaar.

7.2 Vervolgstappen

Gezien de uitkomsten van dit onderzoek en bovenstaande aanbevelingen is het aan te raden om een uitgebreider vervolgonderzoek te doen naar de cognitief/analytische test. Dit in achtneming van bovenstaande aanbevelingspunten, waarbij men kort gezegd de data-analyse uitgebreider zal uitvoeren met als zeer belangrijk punt dat men de beschikking heeft over meer gegevens. Door dit te doen kan er met zekerheid een antwoord gegeven worden op de gestelde hoofdvraag uit dit onderzoek

Naast dit belangrijke punt is het ook aan te raden om de rest van het selectieproces onder de loep te nemen, waarbij gekeken wordt of er in de overige fasen van het proces mogelijk ook een onbedoeld onderscheid wordt gemaakt tussen alloctonen en autoctonen. Hoewel de kans klein is dat er in de rest van het proces onbedoeld onderscheid gemaakt wordt tussen autochtone en allochtone kandidaten, bestaat deze mogelijkheid wel. Uit onderzoek is naar voren gekomen dat beoordelaars tijdens een test of een interview anders kunnen reageren op allochtone kandidaten vergeleken met de reactie op autochtone kandidaten, wat tot een verschil in scores leidt (o.a. Ryan, 2001). Voordat men aan het onderzoeken hiervan begint is het zeer belangrijk dat men hiervoor een goede duidelijke procedure beschrijft om misverstanden te voorkomen.

Als laatste zou de Raad voor de Rechtspraak kunnen overwegen om een andere, bijvoorbeeld minder strenge, aanpak te hanteren voor het selecteren van allochtone kandidaten voor de Raio-opleiding. Er zijn verscheidene zaken waar men aan kan denken. Hierbij moet vermeld worden dat dit zeer riskant kan zijn. Zo kan men er voor kiezen om alloctonen door te laten naar de volgende fase bij een lagere score dan de momenteel gestelde norm (bijvoorbeeld de grens op 450 leggen). Door dit te doen, neemt men echter minder gekwalificeerde kandidaten aan, waardoor de mogelijkheid bestaat dat de kwaliteit van het werk afneemt.

Een andere oplossing voor de minder strenge selectie zou kunnen zijn door het voorleggen van een andere test. Ook zou men alle allochtone kandidaten de uitgebreide versie van de cognitief/analytische test voor kunnen leggen. De kans is echter groot dat hier extra kosten en tijd aan verbonden of zijn. Ook kan het zijn dat een andere test welke men gaat gebruiken, eerst gevalideerd dient te worden voor de groep waarvoor hij gebruikt gaat worden, omdat dit nog niet gedaan is.

In overweging nemend wat er reeds geschreven is over het selecteren van alloctonen en de onderzoeksresultaten, is er weinig wat men kan doen, wil de Raad de kwaliteit van de rechtspraak gelijk houden. Het is aan te raden om rustig af te wachten totdat de samenstelling van de groep vanzelf aantrekt. De verwachting is dat dit in de komende jaren vanzelf gebeurt. Dit is niet wat de Raad graag ziet of doet, maar door dit te doen vermijdt men problemen die zich anders voor kunnen doen bij het afwijken van de vastgestelde normen.

Concluderen kan er gezegd worden dat het selectieproces zoals dat momenteel voor de Raio-opleiding gehanteerd wordt niet in het voordeel van de allochtone kandidaat werkt. Belangrijk is ook dat men zich realiseert dat dit niet nodig is. Het huidige selectieproces is een goed doordacht en goed opgesteld proces, waarbij kandidaten die een grote kans van slagen voor de Raio-opleiding hebben, worden aangenomen en kandidaten met een minder grote kans van slagen niet worden aangenomen. Dit is het algemene doel van een selectieproces (Berry, 2003).

Wat wel belangrijk is in het selectieproces is dat de kandidaten eerlijk op hun 'kunnen' beoordeeld worden. Het moet niet het geval zijn dat gekwalificeerde kandidaten afgewezen worden, doordat er fouten in de test zitten of doordat er tests worden gehanteerd welke om andere redenen niet goed bruikbaar zijn. Om deze reden dienen tests en handleidingen gebruikt te worden voor de selectieprocedure die gevalideerd zijn voor alle bevolkingsgroepen, dus niet alleen voor de meerderheidsgroep, maar ook voor etnische minderheden.

Referenties

- Altink-van den Berg, W.M.M. & Akkerman, A.E. (1989). WIMAS, Vragenlijst voor de Meting van Manipulatief Gedrag in Evers, A., van Vliet-Mulder, J.C. & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland. Deel I Testbeschrijvingen*. Amsterdam: Nederlands Instituut van Psychologen Dienstencentrum B.V.
- Berry, L.M., (2003). *Employee Selection*. Belmont: Wadsworth/Thomson Learning.
- Centraal Bureau voor de Statistiek (2005). Verkregen op 2 mei, 2005, van <http://www.cbs.nl/nl/standaarden/begrippen/bevolking/begrippenlijst.htm>.
- Cohen, E., (2003). *Evaluatie van de Raio-selectieprocedure. Afstudeerverslag in het kader van de opleiding Toegepaste Onderwijskunde aan de Universiteit Twente*. Enschede.
- Drenth, P.J.D. (1965). Test voor Niet-Verbale Abstractie, TNVA in Evers, A., van Vliet-Mulder, J.C. & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland. Deel I Testbeschrijvingen*. Amsterdam: Nederlands Instituut van Psychologen Dienstencentrum B.V.
- Drenth, P.J.D. & Kranendonk, L.J. (1973). Schaal voor Interpersoonlijke Waarden, SIW in Evers, A., van Vliet-Mulder, J.C. & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland. Deel I Testbeschrijvingen*. Amsterdam: Nederlands Instituut van Psychologen Dienstencentrum B.V.
- Drent, P.J.D. & van Wieringen, P.C.W. (1969). Verbale Aanleg Testserie, VAT '69 in Evers, A., van Vliet-Mulder, J.C. & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland. Deel I Testbeschrijvingen*. Amsterdam: Nederlands Instituut van Psychologen Dienstencentrum B.V.
- Ellis, A.P.J. & Ryan, A.M., (2003). Race and Cognitive-Ability Test Performance: The Mediating Effects of Test Preparation, Test-Taking Strategy Use and Self-Efficacy. *Journal of Applied Social Psychology*, 33 (12), 2607-2629.
- Evers, A. & Lucassen, W. (1991). DAT '83 Differentiele Aanleg Testserie in Evers, A., van Vliet-Mulder, J.C. & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland. Deel I Testbeschrijvingen*. Amsterdam: Nederlands Instituut van Psychologen Dienstencentrum B.V.
- Evers, A. & Te Nijenhuis, J., (1999). Liever Speciale dan Traditionele Cognitieve Capaciteitentests voor Allochtonen? Een Vergelijking. *De Psycholoog*, 34 (6), 250-255
- Helms, J.E., (1992). Why is There No Study of Cultural Equivalence in Standardized Cognitive Ability Testing? *American Psychologist*, 47 (9), 1083-1101.
- Helms-Lorenz, M., Van de Vijver, F.J.R & Poortinga, Y.P., (2003). Cross-cultural Differences in Cognitive Performance and Spearman's Hypothesis: g or c? *Intelligence*, 31, 9-29
- Hoekstra, H.A., Ormel, J. & De Fruyt, F., (1996). NEO Persoonlijkheidsvragenlijsten, NEO PI-R en NEO FFI, 1996 in Evers, A., van Vliet-Mulder, J.C. & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland. Deel I Testbeschrijvingen*. Amsterdam: Nederlands Instituut van Psychologen Dienstencentrum B.V.

- Humphreys, L.G., (1986). An Analysis and Evaluation of Test and Item Bias in the Prediction Context. *Journal of Applied Psychology*, 71 (2), 327-333.
- Intelligence Group, (2004). *Kleur! De toekomst van de multiculturele arbeidsmarkt*. Cappelle aan den IJssel: Productie Maximum.
- Miltenburg, E.M., (2005). *Gezocht: Allochtone Rechters; Werven van Etnische Minderheden voor de Raio-opleiding*. Universiteit Twente.
- Murphy, K.R., Cronin, B.E. & Tam, A.P., (2003). Controversy and Consensus Regarding the Use of Cognitive Ability Testing in Organizations. *Journal of Applied Psychology*, 88 (4), 660-671.
- NIP, (2000). *Algemene Standaard Testgebruik*. Amsterdam: Nederlands Instituut van Psychologen.
- Outtz, J.L., (2002). The Role of Cognitive Ability Tests in Employment Selection. *Human Performance*, 15 (1/2), 161-171.
- Ployhart, R.E., Ziegert, J.C. & McFarland, L.A., (2003). Understanding Racial Differences on Cognitive Ability Tests in Selection Contexts: An Integration of Stereotype Threat and Applicant Reactions Research. *Human Performance*, 16 (3), 231-259.
- Raad voor de Rechtspraak, (2003). *Recht doen aan mens en maatschappij. Raio: informatie over de opleiding tot rechter of officier van justitie*. Den Haag
- Raad voor de Rechtspraak, (2005) Verkregen op 28 april, 2005, van <http://www.rechtspraak.nl/Gerechten/RvdR/Over+de+Raad/>
- Raio Info, (2005). Raio-Selectieprocedure.
- Reynolds, C.R., (2000). Why is Psychometric Research on Bias in Mental Testing so Often Ignored? *Psychology, Public, Policy, and Law*, 6 (1), 144-150.
- Sackett, P.R., Schmitt, N., Ellingson, J.E. & Kabin, M.B., (2001). High-Stakes Testing in Employment, Credentialing, and Higher Education. Prospects in a Post-Affirmative-Action World. *American Psychologist*, 56 (4), 302-318.
- Schmidt, F.L., (2002). The Role of General Cognitive Ability and Job Performance: Why There Cannot Be a Debate. *Human Performance*, 15 (1/2), 187-210.
- Schmidt, F.L. & Hunter, J.E., (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 124 (2), 262-274.
- Sloot, B.P., (2004). Moeten Rechters Lijken op de Nederlandse Bevolking? Over de Wenselijkheid van Descriptieve Representatie door de Rechterlijke Macht [elektronische versie]. *Tijdschrift voor de Rechterlijke Macht*, 2, 49-62.
- Tellegen, P., (2000). Verantwoord Testgebruik bij Allochtonen. Een Reactie. *De Psycholoog*, 35 (5), 231-235.
- Te Nijenhuis, J., De Jong, M.J., Evers, E. & Van der Flier, H., (2004). Are Cognitive Differences Between Immigrant and Majority Groups Diminishing? *European Journal of Personality*, 18, 405-434.
- Te Nijenhuis, J. & Evers, A., (2000). De adviezen van Van den Berg en Van Leest (1999) voor Testgebruik bij Allochtonen. Zinnig of Nutteloos? *De Psycholoog*, 35 (4), 169-171.
- Te Nijenhuis, J. & Van der Flier, H., (2003). Immigrant-majority Group Differences in Cognitive Performance: Jensen Effects, Cultural Effects, of both? *Intelligence*, 31, 443-459.

- Van den Berg, R. & Van Leest, P., (1999). Praktisch Testgebruik bij Allochtonen. Wanneer Zinnig, Nuttig of Waardeloos? *De Psycholoog*, 34 (6), 256-260.
- Van den Berg, R. & Van Leest, P., (2000). Het Testen van Allochtonen: Zorgvuldigheid een Voorwaarde. *De Psycholoog*, 35 (4), 172-174.