

# Assessments in selectie- en promotieprocedures: risico's voor ongelijke behandeling

Een literatuurverkenning naar de impact van assessments op inclusie en gelijke kansen



# Inhoudsopgave

Inhoudsopgave	2	Hoofdstuk 2: Intelligentietesten	16
Inleiding	3	2.1 Algemene kritiek op validiteit voorspellend vermogen	16
Onderzoeksvragen	3	2.2 Verschillen tussen mensen met verschillende culturele achtergronden	17
Deelvragen	4	2.3 Verschillen met betrekking tot beperking	18
Onderzoeksmethode	4	2.4 Bang zijn om te voldoen aan een stereotype	18
Opbouw van dit rapport	5	2.5 Dragen intelligentietesten bij aan verschillende kansen in selectieprocessen?	20
Hoofdstuk 1: Persoonlijkheidstesten	6	Samenvatting hoofdstuk 2	20
1.1 Algemene kritiek op validiteit en voorspellend vermogen	6	Hoofdstuk 3: Casuïstiek	21
1.2 In elke culturele context?	8	3.1 De gevonden casussen	21
1.3 Vervalsen van de test?	10	3.2 Wat de casussen laten zien	23
1.4 Mannen als norm?	11	Conclusies en aanbevelingen	24
1.5 Discriminerend naar mensen met een psychische kwetsbaarheid?	12	Deelvragen	24
1.6 Dragen persoonlijkheidstesten bij aan verschillende kansen in selectieprocessen?	13	Hoofdvraag kort beantwoord	24
Samenvatting hoofdstuk 1	14	Aanbevelingen	27
		Bronnen	29

# Inleiding

Bedrijven kunnen gebruik maken van *assessments* om te bepalen of ze een kandidaat wel of niet aannemen en laten doorstromen naar een hogere functie. Volgens uitzendbureau Randstad (z.d.) bestaan assessments vaak uit intelligentietesten en persoonlijkheidsvragen. Er is ook vaak een interview dat hierop toegespitst is, en je krijgt een praktijkcasus voorgelegd. In 2019 concludeerde Nu.nl dat de IQ-test veel voorkomt in sollicitatieprocedures. De nieuwssite onderbouwde dit echter niet met cijfers, maar volgens hoogleraar cross-culturele psychologie Fons van den Vijver worden de testen inderdaad ‘overal in het Westen breed ingezet, van school- en beroepselectie tot promoties’ (Jansen, 2019). Onderdeel van assessments zijn onder meer intelligentietesten, psychologische testen, rollenspellen en interviews, zoals blijkt uit een boek van Bloemers (2013) over assessments. Ook de Nationale Vacaturebank (z.d.) stelt dat assessments vaak bestaan uit een intelligentietest, persoonlijkheidstest (ofwel psychologische test), praktijkopdrachten, rollenspellen en een interview.

In een kennisdocument van de SER Diversiteit in Bedrijf (2023), met betrekking tot werving en selectie, wordt gesteld dat deze assessments problematisch zijn vanuit het oogpunt van inclusie. Op pagina 15 staat: ‘Veel assessments zijn echter ‘gekleurd’ en gebruiken cultuurgebonden

en conventionele selectie-instrumenten. Sollicitanten met een niet-westerse migratieachtergrond zijn daarbij in het nadeel, omdat hun culturele achtergrond niet in de test is meegenomen.’

Via werkgeversorganisatie AWWN kwamen hierover vragen binnen van bedrijven over het gebruik van assessments: is het verstandig om deze in te zetten als je streeft naar een inclusieve organisatie? Anders gezegd: als je wil dat mensen met een migratieachtergrond, maar ook vrouwen (van allerlei culturele achtergronden) en mensen met een beperking goed vertegenwoordigd zijn in alle lagen van je organisatie, is zo’n assessment dan raadzaam? Om deze redenen verrichten we dit verkennende literatuuronderzoek. Hierbij richten we ons op intelligentietesten (ook wel capaciteitstesten genoemd) en persoonlijkheidstesten (ook wel psychologische testen genoemd). Rollenspellen en interviews worden hier buiten beschouwing gelaten om dit verkennende onderzoek af te bakenen.

## Onderzoeksvragen

Onze hoofdvraag: Bestaan er risico’s dat het gebruik van assessments in selectie- en promotieprocedures leidt tot ongelijke behandeling en uitsluiting?

## Deelvragen

- Zijn er risico's als het gaat om gelijke behandeling\* of inclusie\*\*, bij het gebruik van persoonlijkheidstesten in assessments?
- Zijn er risico's als het gaat om gelijke behandeling of inclusie bij het gebruik van intelligentietesten in assessments?
- Zijn er casussen bekend uit de praktijk waarin er risico's naar voren zijn gekomen als het gaat om gelijke behandeling en inclusie met betrekking tot assessments voor selectie en promotie doeleinden?

\* Gelijke behandeling betekent dat alle mensen gelijkwaardig behandeld worden, ongeacht hun geslacht, leeftijd, afkomst, seksuele geaardheid, geloofsovertuiging, handicap of andere persoonlijke kenmerken. Dit houdt in dat er geen discriminatie mag plaatsvinden op basis van deze kenmerken. Gelijke behandeling is daarom het tegenovergestelde van discriminatie.

\*\* Met inclusie of 'insluiting' bedoelen we het doel dat een organisatie kan hebben om te zorgen dat er een representatief personeelsbestand is en een organisatie te zijn waarin iedereen zich thuis voelt en naar vermogen kan werken.

## Onderzoeksmethode

Dit onderzoek betreft een verkennend literatuuronderzoek. We hebben ons allereerst gericht op het verkennen van peer-reviewed literatuur: artikelen gepubliceerd in wetenschappelijke tijdschriften die gecontroleerd zijn door andere onderzoekers. Hiervoor gebruikten we Google Scholar om relevante peer-reviewed wetenschappelijke literatuur te identificeren. Hoewel Google Scholar ook nadelen heeft, wordt het over het algemeen door reviewers als nuttig beoordeeld voor literatuuronderzoek (Falagas et al., 2008; Martín-Martín et al., 2019). Daarnaast hebben we ook gekeken naar bijvoorbeeld proefschriften, mastertheses en rapporten en de 'grijze' literatuur: literatuur die niet peer-reviewed is.

In Google Scholar combineerden we zoektermen als: 'bias' en 'discrimination' met termen zoals 'assessments', 'psychological tests' en 'IQ-test'. Zo hebben we relevante literatuur gevonden, die we vervolgens hebben beschreven in dit rapport. Uiteindelijk vonden we meer dan 80 relevante bronnen die de onderzoeksvraag kunnen beantwoorden. Dit onderzoek is een eerste verkenning en geen systematische review, dus het is mogelijk dat relevante studies ontbreken.

Voor casussen hebben we ons beperkt tot de database van het College voor de Rechten van de Mens, om alleen goed onderzochte en juridisch beoordeelde casussen te presenteren. Het gaat ons echter niet om de juridische vraagstelling, maar om de vraag of assessments bijdragen aan gelijke behandeling en inclusie. Door casussen in kaart te brengen, krijgen we meer zicht op praktijkrisico's.

## Opbouw van dit rapport

In het eerste deel van dit rapport bespreken we de literatuur over persoonlijkheidstesten (ook wel psychologische testen genoemd) en de risico's voor gelijke behandeling en inclusie bij selectie en promotie. In hoofdstuk 2 behandelen we IQ-testen, ook wel intelligentietesten genoemd, en bijkomende risico's. In hoofdstuk 3 presenteren we casussen over (on)gelijke behandeling en inclusie bij selectie en promotie, waarbij gebruik werd gemaakt van persoonlijkheids- en intelligentietesten, uit de database van het College voor de Rechten van de Mens. Tot slot beantwoorden we de onderzoeksvragen in de conclusie van dit rapport.

# Hoofdstuk 1: Persoonlijkheidstesten

In dit eerste hoofdstuk richten we ons op persoonlijkheidstesten die vaak gebruikt worden in assessments. We beschrijven welke mogelijke risico's er zijn volgens de literatuur voor gelijke behandeling en inclusie in organisaties bij selectie en promotie.

## 1.1 Algemene kritiek op validiteit en voorspellend vermogen

Psychologische testen die gebaseerd zijn op de *Big Five* theorie worden gezien als het meest betrouwbaar, omdat dit een gefundeerde wetenschappelijke psychologische theorie betreft (Remann & Nordin, 2021). Met de Big Five worden vijf typen eigenschappen bedoeld:

1. Emotionele stabiliteit (vs. neuroticisme): ervaar je depressie, angst en woede of ben je emotioneel stabiel?
2. Extraversie (vs. introversie): praat je graag en beweeg je makkelijk in sociale situaties of ben je juist stil en teruggetrokken?
3. Openheid voor ervaring (vs. geslotenheid): kies je voor routine of probeer je nieuwe dingen uit?
4. Vriendelijkheid (vs. antagonisme): zoek je de harmonie op of ben je iemand die wrok koestert?
5. Nauwgezetheid (vs. laksheid): ben je ordelijk, heb je zelfdiscipline en ben je betrouwbaar of ben je juist makkelijk afgeleid?

Deze vijf kenmerken zijn ontstaan vanuit analyses van de Engelse taal naar welke woorden worden gebruikt om de persoonlijkheid van anderen uit te drukken (Norman, 1963). Vervolgens zijn er testvragen ontwikkeld bij elk van deze kenmerken om te kunnen achterhalen hoe sterk een individu deze kenmerken bezit (bijvoorbeeld de veelgebruikte *NEO Personality Inventory* van Costa & McCrae, 1992). Daarbij is het gebruikelijk om voor elk kenmerk een aantal onderliggende 'facetten' te meten. Voor vriendelijkheid worden bijvoorbeeld de onderliggende facetten vertrouwen, eerlijkheid, bescheidenheid, samenwerking, sympathie, en altruïsme gemeten. Samengenomen vormen de scores op deze facetten iemands score op het kenmerk vriendelijkheid.

Er is veel wetenschappelijk onderzoek gedaan naar de vraag of persoonlijkheidstesten iets bijdragen aan het verklaren van werkprestaties. Verschillende meta-analyses laten zien dat de persoonlijkheidskenmerken inderdaad positief samenhangen met werkprestaties (Barrick & Mount, 1991; Tett et al., 1991; Zell & Lesick, 2022), verzuim (Ones et al., 2003), en contraproductief werk (Van Iddekinge et al., 2012). Toch zijn hierbij kanttekeningen te plaatsen. Welke persoonlijkheidskenmerken samenhangen met prestaties hangt bijvoorbeeld af van de context van het soort werk dat wordt gedaan (Barrick & Mount, 1991; Tett et al., 1991), maar ook van situationele structuren (duidelijkheid, consistentie, beperkingen en conse-

quenties) en welke kenmerken geactiveerd worden op het werk (Judge & Zapata, 2015). Ook laat onderzoek zien dat met name nauwgezetheid (*conscientiousness*) positief is voor werkprestaties, maar dat de andere persoonlijkheidskenmerken minder sterk of helemaal niet samenhangen met werkprestaties (Zell & Lesick, 2022). Daarnaast laten studies zien dat persoonlijkheidskenmerken doorgaans maar 10 tot 15 procent van de werkprestaties verklaren, en dus 85 procent van de prestaties op werk onverklaard laten (Morgeson et al., 2007).

Niet alleen is de mate waarin iemands persoonlijkheid diens werkprestaties kan verklaren afhankelijk van context factoren, ook hoe iemands persoonlijkheid zich uit of ontwikkelt, is afhankelijk van de context. Zo blijken persoonlijkheidskenmerken bijvoorbeeld af te hangen van het type werk dat men doet. Recent onderzoek laat zien dat mensen in hun werkcontext andere persoonlijkheidskenmerken laten zien dan in andere contexten, omdat hun werk deze kenmerken in hen activeert (Tett et al., 2021). En een studie van Holman en Hughes (2021) liet zien dat de werkcontext er ook voor kan zorgen dat iemand persoonlijkheid verandert, omdat bepaalde kenmerken sterker ontwikkelen als deze nodig zijn voor de baan. De context van een baan is ook erg belangrijk voor hoe goed persoonlijkheidstesten zijn in het voorspellen van prestaties op het werk. Onderzoek van Judge en Zapatta (2015) laat zien dat als een baan specifieke persoonlijkheidskenmerken nodig heeft, dat sommige persoonlijkheidskenmerken een goede voorspellende waarde hebben (bijvoorbeeld extraversie als er voor een baan sociale vaardigheden nodig zijn). Wanneer generiek alle persoonlijkheidskenmerken worden bevraagd, zegt dit dus minder over

hoe iemand op het werk zal presteren dan wanneer er naar specifieke relevante kenmerken wordt gevraagd (Judge & Zapatta, 2015). Het feit dat persoonlijkheidskenmerken context gebonden kunnen zijn, is ook relevant in het kader van de toegenomen digitalisering van assessments (Vervliet, 2022). Want als men deze digitale assessments maakt in een context die niet werkgebonden is (bijvoorbeeld vanuit huis), kan het zo zijn dat persoonlijkheidskenmerken die normaal worden geactiveerd in een werkcontext nu niet geactiveerd worden. Daarmee zou de test dus een minder goede afspiegeling worden van hoe iemands persoonlijkheid zich in die werkcontext uit.

### Objectief werven en selecteren

Iemand toetsen op specifieke competenties of kenmerken die passen bij een specifieke functie – en niet proberen om een totaalbeeld van de persoonlijkheid (of persoonlijkheidskenmerken) van een kandidaat te krijgen – sluit ook aan bij wat genoemd wordt ‘objectief werven en selecteren’. Objectief werven en selecteren betekent dat het gehele proces van werving en selectie wordt geobjectiveerd en op deze manier kan de invloed van (onbewuste) vooroordelen en stereotypen worden verminderd (Felten, 2022; Hulsegge et al., 2020). Bij objectief werven en selecteren, kies je in het begin vier tot zes must haves voor de specifieke functie waarvoor de vacature is en enkel hierop worden de kandidaten beoordeeld en niet op andere eigenschappen of competenties (VIA, 2023).

## 1.2 In elke culturele context?

Culturen zijn niet statisch en altijd in beweging (Kemmelmeier & Kühnen, 2012). En vanuit intersectionele theorie wordt gesteld dat culturele identiteit slechts een van de factoren is die identiteit en positie in de samenleving bepalen en altijd in combinatie met zaken als sekse, leeftijd, seksuele oriëntatie, beperking etc. (Crenshaw, 1989; Stirratt, et al. 2008; Wekker & Lutz, 2001). Desalniettemin is een belangrijk vraag of de persoonlijkheidskenmerken op eenzelfde manier te meten zijn bij mensen met verschillende culturele achtergronden: zijn deze kenmerken universeel of zijn ze in enige mate afhankelijk van de culturele context? Immers, als de kenmerken niet hetzelfde gemeten worden bij mensen met verschillende achtergronden, geeft dat een verkeerd beeld van hun persoonlijkheid en dat kan gevolgen hebben voor selectieprocessen (zie voor review Allik et al., 2023).

Vooropgesteld is een belangrijke kanttekening die gemaakt wordt in het artikel van Laajaj en collega's (2019) dat onderzoeken naar de Big Five persoonlijkheidseigenschappen voornamelijk afgenomen zijn onder wat genoemd wordt 'hoogopgeleide' populaties in landen met een gemiddeld hoog inkomen, in de literatuur ook wel WEIRD-landen genoemd: *Western, Educated, Industrialized, Rich and Democratic*. Zo zijn testen van persoonlijkheidskenmerken in eerste instantie ontwikkeld en onderzocht in Brits en Amerikaans onderzoek, vaak onder witte hoogopgeleide studenten (Costa & McCrae, 1992). Later is onderzocht of deze kenmerken ook getest kunnen worden in andere culturele contexten en onder doelgroepen met andere kenmerken. Hier vindt men wisselende resultaten wat betreft de

validiteit van meetinstrumenten om persoonlijkheid te meten, zoals de *Big Five personality Inventory* (BFI; Laajaj et al., 2019).

Verschillende uitgebreide studies naar persoonlijkheidskenmerken deden onderzoek in meer dan 50 landen, waar studenten werd gevraagd om de persoonlijkheidstesten te doen (McCrae & Terracciano, 2005; Schmitt et al., 2007). Uit deze onderzoeken blijkt dat de vragen in de test ook in andere landen dezelfde vijf persoonlijkheidskenmerken meten. En zo vinden ook Benet-Martinez en John (1998) dat verschillen in culturele waarden niet leiden tot systematische verschillen in de manier waarop de Big Five persoonlijkheidskenmerken co-variëren tussen individuen met een Latijns-Amerikaanse of Anglo-Amerikaanse achtergrond. Dit betekent dat persoonlijkheidstesten in principe een valide manier zijn om persoonlijkheid te meten voor individuen uit verschillende culturele contexten.

Uit verschillende studies blijkt echter ook dat de gemiddelde scores op de Big Five verschillen per land. Daaruit blijkt dat landen die geografisch dicht bij elkaar liggen vaak ook een vergelijkbaar gemiddeld persoonlijkheidsprofiel hebben. Ook laten deze studies zien dat gemiddelde persoonlijkheidsprofielen in wat in deze studies genoemd wordt 'Amerikaanse culturen' en 'Europese culturen' verschillen van die in 'Aziatische culturen' en 'Afrikaanse culturen' (Allik & McCrae, 2004; Schmitt et al., 2007). Alhoewel culturen niet statisch zijn en er in één land uiteraard mensen wonen met uiteenlopende culturele achtergronden, lijken er toch enige verschillen te zijn in hoe mensen uit verschillende landen gemiddeld gezien scoren op bijvoorbeeld extraversie of openheid.



Toch zijn er verschillende punten van kritiek op de vergelijkbaarheid van persoonlijkheidskenmerken in verschillende culturele contexten. De testen zijn niet altijd in iedere context goed bruikbaar aangezien zoals hierboven genoemd is, ze afkomstig zijn uit een witte Westerse context van ‘hoogopgeleiden’. Onderzoeken die zijn gedaan binnen andere sociale contexten – anders dan de Westerse, witte context – laten zien dat ook andere persoonseigenschappen belangrijk kunnen worden gevonden, zo blijkt uit een recente literatuurstudie van Sitinjak (2023). Denk bijvoorbeeld aan openheid: dat kan in bepaalde contexten minder relevant worden gevonden. En in bepaalde culturele contexten worden bijvoorbeeld eerlijkheid en nederigheid juist belangrijk gevonden (Sitinjak, 2023). De Big Five kenmerken geven mogelijk dus een te beperkt beeld van persoonlijkheid. In een studie bleek er bijvoorbeeld nog een zesde kenmerk zijn beschreven, namelijk die van eerlijkheid (vs. bescheidenheid) (Ashton et al., 2004). In China bleek er nog het kenmerk ‘Interpersoonlijke verbondenheid’ te ontwaren (Cheung et al. 2001) en dit kenmerk kon vervolgens ook gemeten worden bij Amerikaanse respondenten, dus bleek óók buiten China relevant (Cheung et al., 2003). Er zouden dus meer kenmerken van persoonlijkheid gemeten kunnen worden dan de Big Five en een witte Westerse blik op persoonlijkheid is mogelijk wat te beperkend.

Toch zijn er verschillende punten van kritiek op de vergelijkbaarheid van persoonlijkheidskenmerken in verschillende culturele contexten. Juist omdat de Big Five persoonlijkheidskenmerken getoetst zijn onder voornamelijk witte, theoretisch opgeleide mensen, kennen dergelijke assessments een sterke bias met betrekking tot wat er in de westerse

witte context wordt gezien als belangrijke persoonlijkheidskenmerken. De testen zijn daardoor niet altijd in iedere (culturele) context goed bruikbaar. Zo zijn er studies die stellen dat de Big Five kenmerken een te beperkt beeld geven van persoonlijkheid en dat er in andere culturele contexten ook meer dan deze vijf kenmerken van belang zijn voor het beschrijven van iemands persoonlijkheid (Sitinjak, 2023). Onderzoeken die zijn gedaan binnen andere sociale contexten – anders dan de westerse, witte context – laten zien dat ook andere persoonseigenschappen belangrijk zijn om te kunnen meten middels een assessment. Enkele voorbeelden daarvan zijn: ‘Eerlijkheid (vs. bescheidenheid)’ (Ashton et al., 2004) en ‘Interpersoonlijke verbondenheid’ te ontwaren (Cheung et al. 2001; Cheung et al., 2003).

Bovendien is in sommige onderzoeken ook geen empirisch bewijs te vinden voor de Big Five persoonlijkheidskenmerken. In het onderzoek van Laajaj en collega’s (2019) vond men bijvoorbeeld geen vijf factor structuur van persoonlijkheid waarmee de Big Five normaliter wordt aangeduid in een groot deel van de persoonlijkheidstesten die werden afgenomen in landen met een gemiddeld laag of gemiddeld inkomen. Ook Gurven en collega’s (2013) laten zien dat de Big Five niet gevonden werd onder de Tsimane, een etnische groep uit Bolivia.

Een ander, meer fundamenteel, kritiekpunt is dat het concept persoonlijkheid niet in alle contexten hetzelfde wordt gebruikt. In een kritisch overzichtsartikel van Carlo en collega’s (2014) wordt benoemd dat persoonlijkheid vanuit de psychologie wordt gezien als iets wat cultuur-overstijgend is, maar waarbij verschillende expressies of manifestaties

van persoonlijkheid zijn in verschillende culturen. Sommige cultureel-anthropologen, daarentegen, begonnen vanaf de jaren '70 persoonlijkheid te zien als iets wat geworteld is in cultuur omdat culturele processen iemands *sense of self* definiëren. Ook empirische studies laten zien dat er in verschillende culturele contexten anders over persoonlijkheid wordt gedacht. Zo laat onderzoek zien dat in culturele contexten waar individualisme belangrijk is mensen over hun persoonlijkheid denken als een vaststaand kenmerk van henzelf, dat weinig verandert over de tijd of in verschillende sociale contexten. Terwijl mensen in culturele contexten waar collectivisme belangrijk is, persoonlijkheid als meer fluïde zien, waarbij het gepast is om kenmerken aan te passen aan de sociale context (Gage et al., 2015; Church et al., 2012). Bovendien kunnen persoonlijkheidskenmerken niet alleen afhangen van de sociale context, maar ook van taal of het werk dat men doet. Zo zijn er studies die aantonen dat personen die zich identificeren met meerdere culturen andere resultaten laten zien bij persoonlijkheidstesten wanneer ze de vragen krijgen aangeboden in de taal van de ene of de andere cultuur (Chen & Bond, 2010; Lönnqvist et al., 2014). Uit onderzoek van KIS is tevens bekend dat jongeren met een migratieachtergrond vaak schakelen en navigeren tussen verschillende leefwerelden en zich via trial en error leren verhouden tot verschillende sociale domeinen zoals thuis, school en werk (Day en Badou, 2019). Mensen met een migratieachtergrond passen zich dus vaak aan door de jaren heen, aan wat de verschillende contexten van hen vragen (Day en Badou, 2019). Het vaststellen van één type persoonlijkheid lijkt dus mogelijk niet goed te passen bij de realiteit waarin mensen zich aanpassen aan verschillende contexten en dat geldt mogelijk in het bijzonder voor mensen die opgroeien met

meerdere culturele referentiekaders. Ook internationaal onderzoek laat zien dat mensen die uit een ander land komen vaak meerdere culturele identiteiten ontwikkelen (Chen et al., 2008) en dat zij in sommige situaties als het ware switchen tussen verschillende culturele identiteiten in verschillende contexten zoals tussen thuis en werk (Schwartz, et al, 2019). Dat leren mensen waarschijnlijk al jong: een studie onder kinderen van Griekse afkomst in Nederland laat zien dat zij al switchen tussen verschillende culturele raamwerken (Verkuyten & Pouliasi, 2002).

Kortom: persoonlijkheidskenmerken kunnen voor iedereen verschillen per context, waardoor dit mogelijk verschilt tussen bijvoorbeeld een werk en privé context. Maar voor respondenten die opgegroeid zijn met meerdere culturele referentiekaders kunnen de resultaten van de persoonlijkheidstest nog meer variëren afhankelijk van de context en de taal, en daarmee ook een minder valide meetinstrument zijn.

### 1.3 Vervalsen van de test?

Wanneer vacatureteksten er blijk van geven dat er iemand met een bepaalde persoonlijkheid wordt gezocht, zullen mensen hier hun antwoorden op aanpassen om daardoor beter bij de vacature te passen. Studies laten zien dat ongeveer één derde tot de helft van sollicitanten hun antwoorden op een persoonlijkheidstest verdraaien of aanpassen om beter bij de vacature te passen (Donovan et al. 2003; Griffith et al., 2007). Soms wordt er daarom ook getoetst hoe bereidwillig men is om sociaal wenselijk te antwoorden. Maar onderzoekers hebben geconcludeerd dat deze sociale wenselijkheid

geen goede manier is te controleren voor de effecten van vervalsen weg te nemen (Ellingson et al., 1999). Het is dus lastig om het vervalsen van persoonlijkheidstesten te ondervangen. En ook als persoonlijkheidstesten worden vervalst door respondenten, blijken de testen toch nog een redelijke voorspeller te zijn van prestaties op het werk (Ones, Viswesvaran & Reiss, 1996). Vandaar dat persoonlijkheidstesten toch nog breed worden ingezet.

Maar vervalsingen op persoonlijkheidstest zijn wel van invloed op de selectie van kandidaten: kandidaten die hun persoonlijkheidstest hebben vervalst (en daarvoor niet wordt gecompenseerd) worden hoger gewaardeerd en eerder geselecteerd in selectieprocedures (Komar, Brown, Komar, & Robie, 2008). Dat is ook van belang in het kader van culturele achtergrond, omdat de mate waarin mensen hun persoonlijkheidstesten aanpassen gemiddeld verschilt tussen culturele groepen. Een studie toonde aan dat respondenten met een Aziatisch Amerikaanse, of Zwart Amerikaanse achtergrond hun antwoorden op de persoonlijkheidstest minder aanpasten aan de vacaturetest dan dat witte Amerikanen deden (McGee & McGee, 2022). Dat zou ervoor kunnen zorgen dat witte Amerikanen in het voordeel zijn bij selectieprocedures. Een andere studie (Ones et al. 1996) liet juist zien dat minderheden meer sociaal wenselijke antwoorden invulden dan witte Amerikanen. Deze studie liet ook zien dat het controleren voor sociale wenselijkheid ervoor kan zorgen dat minderheden in het nadeel zijn, omdat deze gecontroleerde scores een vertekend beeld geven. Sociale wenselijkheid zou je in deze gevallen namelijk ook

kunnen zien als een aanpassend vermogen, en daarmee zelfs wenselijk kunnen zijn voor de werkcontext (Tett & Simonet, 2021).

Vanuit KIS hebben wij er in dit rapport voor gekozen om Zwart met een hoofdletter te schrijven. Dit om te verwijzen naar mensen met een gedeelde transnationale politiek en culturele identiteit die door mensen in de Afrikaanse diaspora gedeeld worden. Zwart refereert dus niet direct (en uitsluitend) naar de kleur, maar gaat verder dan dit. Het refereert ook naar mensen die een gedeelde geschiedenis, identiteit en gemeenschap dragen (zie [Zwart Manifest](#)). Als er in Europa naar 'witte mensen' wordt verwezen, wordt er veelal niet verwezen naar een gedeelde identiteit en culturele geschiedenis. Wanneer dit wel wordt gedaan spreken we (vaak) over een specifieke groep, denk aan Nederlanders, Fransen etc., Daarom hebben wij ervoor gekozen wit met een kleine letter te schrijven. KIS sluit hiermee aan bij invloedrijke media zoals de New York Times (zie [Why We're Capitalizing Black - The New York Times](#) (nytimes.com)) en AP News ([AP changes writing style to capitalize "b" in Black](#) | AP News).

## 1.4 Mannen als norm?

Als het gaat om gender en persoonlijkheidstesten, dan wordt in een studie van Stabile (2001) onderbouwd dat deze testen mensen die getalsmatig in de minderheid zijn, kunnen benadelen, doordat ze vaak getest worden onder groepen mensen die niet representatief zijn voor de gehele bevolking. In het artikel van Kubiak en collega's (2023) wordt beschreven dat

er nog veel discussie is tussen onderzoekers over genderverschillen in persoonlijkheid. Sommige onderzoekers suggereren dat er voornamelijk genderverschillen te vinden zijn in persoonlijkheidskenmerken, zoals vriendelijkheid en emotionele stabiliteit/neuroticisme. Zo wordt in het artikel van Schmitt en collega's (2017) genoemd dat mannen relatief lager scoren dan vrouwen op neuroticisme en vriendelijkheid, en in mindere mate op extravertie en openheid voor ervaring. Maar andere empirische studies laten zien dat de genderverschillen in persoonlijkheid tussen mannen en vrouwen klein zijn en dat er vooral veel gelijkheid is (Zell et al., 2015). Recent onderzoek toont tevens aan dat gendergelijkenissen en verschillen mede afhankelijk zijn van hoe de data wordt georganiseerd. Wanneer er wordt gekeken naar de onderliggende facetten van een persoonlijkheidskenmerk worden er bijvoorbeeld vaker genderverschillen gevonden. Maar als diezelfde scores op de facetten worden samengenomen om het persoonlijkheidskenmerk te meten, dan zijn er nauwelijks genderverschillen te meten (Eagly and Revelle, 2022).

Hoe sterk genderverschillen zijn, blijkt volgens enkele onderzoeken afhankelijk te zijn van de culturele context: in landen waar meer gelijkheid is tussen vrouwen en mannen (bijvoorbeeld egalitaire genderrollen, gender socialisatie en sociaalpolitieke gendergelijkheid), zijn de persoonlijkheidsverschillen tussen vrouwen en mannen, zoals gemeten op de Big Five, over het algemeen groter (Schmitt et al., 2008; 2017; Giolla & Kajonius, 2019). Echter geven Schmitt en collega's (2017) aan dat dergelijke factoren de interpretatie van persoonlijkheid-gerelateerde gedragingen door mannen en vrouwen kunnen beïnvloeden: bepaalde gedragingen

in culturen waar individualisme gemiddeld gezien belangrijk is, worden vaker toegeschreven aan iemands 'ware' persoonlijkheid en in culturen waar collectivisme gemiddeld gezien, worden vaak toegeschreven aan de genderrollen van mannen en vrouwen. Een voorbeeld hiervan is dat vriendelijk gedrag uitgedragen door een man of vrouw in een meer egalitaire (en individualistische) cultuur gezien kan worden als een vrijwillige en gekozen gedraging, en daarom een reflectie is van de 'ware' persoonlijkheid van een persoon. Op deze manieren worden de genderverschillen in Noord-Europese culturen gezien als 'echte' gender verschillen, terwijl vriendelijk gedrag door een persoon uit een minder egalitaire (en collectivistische) cultuur sneller gezien wordt als een reflectie van normen in genderrollen voor vrouwen (en in dit geval niet voor mannen) (Schmitt et al., 2017). Het afwezig zijn van genderverschillen in persoonlijkheid in minder egalitaire en meer collectivistische culturen wordt in deze dus toegeschreven aan genderrollen in plaats van persoonlijkheidskenmerken van personen.

## 1.5 Discriminerend naar mensen met een psychische kwetsbaarheid?

Persoonlijkheidstesten gebaseerd op de Big Five worden gezien als valide, maar zoals Timmons (2021) beschrijft, voorspelt de Big Five ook neurodivergentie.

Zie [Neurodivergentie en neurodiversiteit: wat betekent het? - IDEM Rotterdam](#) voor meer uitleg.

Zo is bekend uit een recent longitudinaal onderzoek dat er een sterke link is tussen hoe je scoort op een persoonlijkheidstest gebaseerd op de Big Five en het hebben van ADHD (Bell et al., 2024). Ook een eerdere studie toonde al een verband aan tussen het hebben van ADHD en de persoonlijkheidskenmerken van de Big Five (Nigg et al., 2002). Zo scoren mensen met ADHD minder hoog op *conscientiousness* (Bell et al., 2024; Nigg et al., 2002), wat onder meer in beeld brengt of je snel afgeleid bent of juist ordelijk (Timmons, 2021). En ook autisme is gerelateerd aan de eigenschappen die gemeten worden met de Big Five. Op die manier legt een test op basis van de Big Five een aanleg voor autisme bloot, zo laat een cross-sectionele analyse onder honderden mensen uit verschillende werelddelen zien (Robinson et al., 2020). Dit komt tevens voort uit een meta-analyse: zo scoren mensen met autisme significant lager op *agreeableness*, *openness*, *conscientiousness* en *extraversion*, en scoren ze juist lager op emotionele stabiliteit (en dus hoger op neuroticisme; Lodi-Smith et al., 2019). Dat betekent, zoals Timmons (2021) beargumenteert, dat persoonlijkheidstesten die worden gebaseerd op de Big Five (en in de Verenigde Staten is dat de meerderheid) de werkgever zicht geven op wie wel en wie niet neurodivergent is en wie afwijkt van de neurotypisch norm. De bedoeling is dat deze test mede bepaalt wie wel of niet de baan krijgt, en daarmee werkt deze discriminatie van mensen met een psychische kwetsbaarheid in de hand.

## 1.6 Dragen persoonlijkheidstesten bij aan verschillende kansen in selectieprocessen?

Persoonlijkheidskenmerken kunnen een rol spelen bij selectieprocedures. Met betrekking tot culturele achtergrond zijn de studies niet eenduidig. Sommige studies vinden wel verschillende kansen in selectie door persoonlijkheidstesten (Risavy & Hausdorf, 2011), terwijl anderen stellen dat de verschillen verwaarloosbaar zijn (Ones & Anderson, 2002). Als het gaat om neurodiversiteit is er een voorbeeld gevonden in de literatuur waaruit blijkt dat het gebruiken van zo'n test kan leiden tot discriminatie. In de Verenigde Staten heeft de rechter beoordeeld dat een persoonlijkheidstest, de *Minnesota Multiphasic Personality Inventory* (MMPI), een vorm van een medische test is, omdat het de mogelijkheid heeft om psychische problematiek op te sporen (Timmons, 2021). Dat is in strijd met de wet tegen discriminatie jegens personen met een beperking. Deze MMPI-2 sluit tevens aan bij de Big Five (Harkness, McNulty, Ben-Porath, 1995). Persoonlijkheidstesten kunnen dus een *bias* bevatten en aanwakkeren ten aanzien van mensen die neurodivergent zijn, zo wordt beargumenteerd door verschillende onderzoekers (Stabile, 2001; Wiggleton-Little & Callender, 2023). Een van de argumenten is dat door het gebruik van persoonlijkheidstesten je de autistische kandidaten er dus kunt 'uithalen': dat betekent dat deze testen het discrimineren van deze groep makkelijker maken omdat nu duidelijk is wie zich wel of niet op het spectrum bevindt (Wiggleton-Little & Callender, 2023).

Ook op het gebied van gender kunnen er verschillen in het selectieproces ontstaan. Onderzoek van Powell en collega's (2011) laat bijvoorbeeld

zien dat er inderdaad kleine verschillen kunnen ontstaan in de selectieprocedures waarin deze persoonlijkheidsassessments worden gebruikt: vrouwen worden dan net iets minder vaak geselecteerd dan mannen gebaseerd op hun persoonlijkheidsscores op de kenmerken extravertie en nauwgezetheid. Maar dat is vooral het geval als er een uitgebreide persoonlijkheidstest wordt gedaan, waarbij ook wordt gekeken naar de onderliggende facetten van elk kenmerk. Daarbij bleek dat een aantal van die facetten (zoals *agency* en *communion*) vooral de verschillen veroorzaakte. Wanneer alleen op de bovenliggende Big Five kenmerken werd getoetst, waren er nauwelijks genderverschillen in selectie (Powell et al. 2011). Ook recent onderzoek van Kubiak en collega's (2023) onderschrijft deze bevinding: wanneer niet de facetten worden bekeken, maar de bovenliggende kenmerken zijn er nauwelijks genderverschillen in selectieprocedures. Omdat persoonlijkheid, op deze manier gemeten, weinig verschilt tussen mannen en vrouwen, zien deze onderzoekers persoonlijkheidstesten ook juist als een manier om genderbias in werving en selectie tegen te gaan.

Hierbij maakt het wel uit hoe je de persoonlijkheidstesten gebruikt in selectieprocedures. Onderzoek laat zien dat als persoonlijkheidstesten worden gebruikt op een manier waarbij iedereen die niet hoog genoeg scoort op bepaalde facetten niet doorgaat in de selectieprocedure, dat sollicitanten met een zogenaamde minderhedenachtergrond dan minder vaak worden geselecteerd. Maar als er andere methoden worden gebruikt, door bijvoorbeeld niet naar facetten te kijken, of gewogen scores te gebruiken, dan zijn er geen negatieve gevolgen voor minderheden (Risavy & Hausdorf, 2011).

Persoonlijkheidstesten zouden, volgens sommigen, ook kunnen worden gebruikt in het voordeel van sollicitanten die getalsmatig in de minderheid zijn op de werkvloer. Zo laat een studie van Newman en Lyon (2009) zien dat wanneer men enkel aan sollicitanten van kleur ook vraagt om een cognitieve toets en test om hun nauwgezetheid te meten, dat deze combinatie van toetsen eraan kan bijdragen dat sollicitanten met een van kleur vaker worden geselecteerd. Dat vraagt echter om een geheel andere inzet dan hoe de testen nu vaak worden gebruikt.

## Samenvatting hoofdstuk 1

Persoonlijkheidstesten kunnen in enige mate een voorspelling doen van hoe iemand gaat presteren op het werk. Hoe sterk persoonlijkheid die prestatie kan voorspellen, hangt uiteindelijk ook sterk af van welke kenmerken passen bij welk type werk. Niet alle persoonlijkheidskenmerken zijn dus even relevant voor elke baan. De waarde van zo'n psychologische test is mogelijk ook minder omdat deze test kunnen worden vervalst door de antwoorden zo in te vullen dat deze passen bij wat er wordt gezocht. Het is moeilijk is om te controleren voor deze sociale wenselijkheid, zeker in culturele contexten waar het maken van dergelijke aanpassingen als gewenst gedrag wordt gezien. Vanuit het perspectief van gendergelijkheid zijn er enige kritische kanttekeningen te maken, omdat vrouwen en mannen anders kunnen scoren op testen. Dit hangt ook af van cultuur hoe sterk deze verschillen zijn. Want hoewel er bewijs is dat de veel gebruikte testen, gebaseerd op de psychologische theorie van de The Big Five kenmerken, in verschillende culturen bruikbaar is, is er ook bewijs van het tegendeel. Zo

worden gemiddeld gezien bepaalde eigenschappen anders gewaardeerd in verschillende culturele contexten en is persoonlijkheid onderr andere. ook afhankelijk van de culturele context en dus niet altijd eenduidig en statisch. Daarnaast is er veel kritiek op deze persoonlijkheidstesten – in het bijzonder de veelgebruikte Big Five – omdat het werkgevers inzicht geeft in wie wel of niet kan worden gezien als neurodivergent. Een werkgever die bijvoorbeeld niet de baan wil geven aan iemand die autistisch is, ADHD heeft of een angststoornis, kan op basis van deze test deze mensen als dusdanig herkennen. Persoonlijkheidstesten kunnen dus bijdragen aan ongelijke behandeling van neurodivergente personen, mensen met een migratieachtergrond en vrouwen. Het is belangrijk om tijdens selectie te overwegen of een dergelijke test echt nodig is. Mocht dat het geval zijn, is het zaak goed na te denken over hoe deze test op een objectieve manier wordt ingezet: welke aspecten van persoonlijkheid echt relevant zijn voor de baan? Worden deze dan ook in een werk-gerelateerde context bevraagd? Welke methode wordt gebruik voor hoe de testcores leiden tot selectie van een kandidaat?

# Hoofdstuk 2: Intelligentietesten

**In dit hoofdstuk gaan we in op de intelligentietesten (ook wel capaciteitstesten genoemd) en specifiek IQ-testen. We beschrijven of en, zo ja, welke aanwijzingen er zijn vanuit de literatuur dat deze testen mensen kunnen benadelen en dus mogelijk haaks kunnen staan op het streven naar gelijke behandeling en inclusie. Hierbij kijken we naar gender, mensen met een migratieachtergrond en mensen met een beperking.**

## 2.1 Algemene kritiek op validiteit voorspellend vermogen

Uit verschillende studies komt naar voren dat de score van IQ-testen een van de voorspellers is van academische prestaties (Heaven & Ciarrochi, 2012; Sternberg, Grigorenko & Bundy, 2001). Het is echter wel slechts één van de voorspellers want er blijken nog andere voorspellers; zo is zelfcontrole van invloed (Duckworth & Seligman, 2005; Vazsonyi, Javakhishvili & Blatny, 2022). Maar dat neemt niet weg dat een IQ van enige voorspellende waarde is. Maar meet een IQ-test dan ook wel echt intelligentie? Op dat idee is veel kritiek. Econometrist en De Correspondent-redacteur Sanne Blauw wijst erop dat intelligentie een abstract concept is dat we zijn gaan meten (Jansen, 2019). Een IQ-test meet voornamelijk zaken die te maken hebben met abstract denken, zoals cijferreeksen, ruimtelijk inzicht en

vocabulaire. De focus die wordt gelegd op abstract denken is volgens haar geen objectieve keuze, maar een waardeoordeel. Ze stelt dat er in principe weinig mis is met de test zolang men voor ogen heeft dat een IQ-score niet synoniem is aan intelligentie (Jansen, 2019). Ook volgens een ander artikel in De Correspondent is intelligentie volgens prominente psychologen meer dan dat wat gemeten kan worden op een IQ-test (Eerkens, 2015). Volgens een academisch artikel hierover is het IQ meer een weging van specifieke cognitieve vaardigheden dan de reflectie van een onderliggend algemeen cognitief functioneren (Kovacs & Conway, 2019). In een review uit 2005 wordt geconcludeerd door Nettelbeck en Wilson dat er sterk bewijs is dat intelligentie enerzijds en IQ anderzijds andere constructen zijn, maar dat een IQ-test tot nu toe het beste middel is om iets te zeggen over intelligentie.

Intelligentietesten worden nog steeds gezien als belangrijke indicatoren voor werkprestaties. En er zijn ook veel studies die laten zien dat intelligentietesten redelijk zijn in het voorspellen van werkprestaties. Ook een recente en zeer uitgebreide meta-analyse (Sackett, et al. 2022) laat zien dat intelligentietesten een goede voorspellende waarde hebben. Maar daar waar eerder onderzoek intelligentietesten als een van de belangrijkste en meest voorspellende factoren liet zien, met een voorspellende waarde van meer dan 50 procent (bijvoorbeeld Schmitt & Hunter, 1998), laat deze nieuwe studie zien dat deze voorspellende waarde zeer



waarschijnlijk te groot wordt ingeschat door onderliggende statistische aannames. In de studie van Sackett en collega's (2024) wordt hier beter rekening mee gehouden en is de voorspellende waarde van algemene intelligentietesten geschat op 31 procent. Bovendien bleken er andere selectiemethoden beter te presteren in het voorspellen van werkprestaties. Gestructureerde interviews waren het meest voorspellend (42 procent) gevolgd door vakkennis test (40 procent). Opvallend daarbij is dat in deze verbeterde meta-analyse er een veel duidelijkere voorspellende waarde uitgaat van assessments die gaan over de specifieke werkcontext, en niet zozeer om algemene vaardigheden (of persoonlijkheid), zoals intelligentietesten. Omdat intelligentietesten wellicht dus minder voorspellend zijn, wordt vaak aangeraden om, als deze toetsen worden gebruikt, dit te doen in combinatie met andere selectie instrumenten zoals interviews of een test van vakkennis (Ployhart & Holtz, 2008; Sackett et al., 2022). Sterker nog, recent werk (Berry et al. 2024) laat ook zien dat als er andere methoden, zoals gestructureerd interviewen, worden gebruikt, het gebruik van algemene intelligentietesten niet van toegevoegde waarde is om werkprestaties te verklaren. Het weglaten van intelligentietesten zorgde er niet voor dat werkgevers slechter konden voorspellen hoe een kandidaat zou presteren op het werk.

## 2.2 Verschillen tussen mensen met verschillende culturele achtergronden

Er zijn onder meer aanwijzingen dat IQ-testen die kunnen worden gebruikt in assessments niet goed bruikbaar zijn in etnisch diverse omgevingen

(Bloemers, 2013). In een studie van Shuttleworth-Edwards (2016) in Zuid-Afrika worden twee type intelligentietesten – de *Wechsler Adult Intelligence Scale* versie 3 en 4 (WAIS-III en WAIS-IV) – kritisch geëvalueerd en is de conclusie dat er een tekortkoming is in beide testen in het onvermogen om onderscheid te maken tussen Afrikaanse moedertaalsprekers met een achtergrond van bevoorrecht onderwijs en degenen die uit onderwijsachterstandssituaties komen. In een andere studie naar verschillen in scores op de Wechsler Abbreviated Scale of Intelligence (WASI) tussen vloeiend Engelsprekende, etnisch diverse mensen (met een migratieachtergrond) en eentalige Engelsprekende Anglo-Amerikanen, blijkt dat de tweede groep beter presteert dan de eerste groep op de verbale sub-testen, maar niet op de non-verbale sub-testen (Razani et al., 2007). Hierop aansluitend: een recente meta-analyse in het Verenigd Koninkrijk geeft aan dat mensen met verschillende culturele achtergronden wel degelijk verschillend scoren; doorgaans zijn deze voor de mensen met een migratieachtergrond lager dan voor de witte meerderheidsgroep (met als uitzondering enkele Aziatische groepen) (te Nijenhuis et al. 2024).

Er zijn echter ook andere onderzoeken die laten zien dat er geen culturele verschillen zijn bij het afnemen van intelligentietesten, en dat veel gestandaardiseerde testen een valide manier zijn van het meten van intelligentie ongeacht culturele achtergrond (Reynolds et al., 2021).

Er zijn ook verschillende onderzoeken gedaan in Nederland. Duidelijk is dat mensen met een migratieachtergrond in Nederland gemiddeld lager scoren op IQ-testen (te Nijenhuis et al., 2004). Hoe dat precies komt, ligt complex.

Bekend is uit onderzoek dat mensen met een migratieachtergrond gemiddeld hoger scoren op kortetermijngeheugen, maar gemiddeld lager op ruimtelijke visualisatie, maar onduidelijk is waarom dat zo is. Erfelijkheid van intelligentie (overgekregen van ouders) wordt soms ook genoemd als mogelijke verklaring, maar recent onderzoek laat zien dat erfelijkheid voor alle culturele groepen een even grote rol speelt voor intelligentie en dus geen verklaring kan zijn voor onderliggende verschillen (Pesta et al. 2020). Daarnaast liggen verklaringen onder meer in *language bias*: als Nederlands niet je moedertaal is, scoor je lager op de test, zo wordt geconcludeerd in het onderzoek van te Nijenhuis en collega's (2016). Onduidelijk blijft echter of deze verschillen tussen mensen met en zonder migratieachtergrond blijven bestaan, wanneer er wordt gekeken naar mensen met een migratieachtergrond voor wie Nederlands wel de moedertaal (of een daarvan) is.

Naast het verschil in moedertaal, speelt ook onderwijs een rol. Sowieso blijkt onderwijs een belangrijke manier om intelligentie te vergroten. Uit een meta-analyse in de Verenigde Staten blijkt dat elk jaar onderwijs 1 tot 5 punten aan intelligentie kan verhogen (Ritchie & Tucker-Drob, 2018). Maar voor migranten lijkt een verklaring ook mogelijk te liggen in het aantal jaren van onderwijs in het betreffende land waar de test wordt afgenomen en of iemand eerste of tweede generatie migrant is, zo blijkt uit Amerikaans onderzoek (Hausdorf & Robie, 2018). Dit komt overeen met onderzoek in Nederland dat ook laat zien dat eerste generatie migranten gemiddeld lager scoren dan tweede generatie (te Nijenhuis et al., 2004). Duidelijk is dus in ieder geval dat IQ-testen mensen benadelen die niet Nederlands als moedertaal hebben en niet in Nederland naar school zijn geweest.

## 2.3 Verschillen met betrekking tot beperking

Mensen die dyslectisch zijn, scoren vaker minder hoog op IQ-testen, zo wordt onderbouwd in een Nederlandstalig artikel (Ronner & Noks Nauta, 2010). IQ-testen worden ook ingezet om dyslectisch aan te tonen, maar dat blijkt ook niet geheel betrouwbaar (Toth & Siegel, 2020).

Ook dyscalculie kan de uitkomsten van een IQ-test negatief beïnvloeden: er wordt dan vaak vanuit gegaan dat zwakke rekenscores het gevolg zijn van een laag IQ, maar tegelijkertijd is het zo dat mensen met een rekenstoornis lager zullen scoren op een intelligentietest, aangezien die testen een beroep kunnen doen op rekenvaardigheden (De Kock, Desoete, Stock, 2008). Er is ook onderzoek dat aantoont dat kinderen met ADHD lager scoren op IQ-testen (Brown, Reichel, Quinlan, 2009) en dit geldt ook voor een deel van de volwassenen met ADHD (Bridgett & Walker, 2006).

## 2.4 Bang zijn om te voldoen aan een stereotype

Stereotypen verwijzen in de psychologie naar de cognitieve denkbeelden die mensen hebben over bepaalde groepen, en dan in het bijzonder naar de overdreven denkbeelden over groepen mensen (Felten & Broekroelofs, 2022; Gordijn & Wigboldus, 2013). Mensen die onderdeel zijn van de groepen die gestereotypeerd worden, zijn vaak bang om te voldoen aan deze stereotiepe beelden en dit wordt stereotype threat genoemd (Spencer, Logel & Davies, 2016). Dit kan negatieve gevolgen hebben voor prestaties. Uit een meta-analyse in 2008 blijkt dat wanneer vrouwen en/of mensen van kleur bewust of onbewust een herinnering krijgen van een stereotype

beeld van hun groep, beide groepen slechter presteren op verschillende soorten cognitieve vaardigheidstesten in vergelijking met degene die niet zo'n herinnering krijgen (Nguyen & Ryan, 2008). Dit lijkt mogelijk ook te gelden voor mensen met leerproblemen zoals dyslexie (Haft, Greiner de Magalhães & Hoeft, 2023).

Zo'n 'herinnering' van een stereotiep beeld is bijvoorbeeld vertellen dat vrouwen slechter zijn in wiskunde of dat mensen van kleur slechter scoren op de test. Wanneer je dit leest of hoort, kun je dus bang worden om aan dit stereotiep beeld te voldoen, waardoor prestaties achteruitgaan (Nguyen & Ryan, 2008). Dit speelt niet alleen in de Verenigde Staten, maar ook in Europa: zie bijvoorbeeld een recente studie uit Duitsland (Froehlich et al., 2022) en een meta-analyse uit 2015 van Appel en collega's. Maar zo'n herinnering aan een stereotiep beeld kan heel subtiel zijn, zo blijkt uit experimenteel onderzoek van Steele en Aronson (1995): het kan gaan om je afkomst invullen op een formulier of om expliciet te maken dat het gaat om een test om je intelligentie te meten. Want zoals aangetoond is in dit experimentele onderzoek, leidt dit tot angst om te voldoen aan het stereotype beeld minder intelligent te zijn en dat leidt tot slechtere testprestaties (Steele & Aronson, 1995). Dit maakt het aannemelijk dat mensen met een migratieachtergrond die zich bewust zijn van dat er stereotype beelden over hun groep zijn als 'minder intelligent', onderpresteren op IQ-testen in assessments op bijvoorbeeld het werk (Walton et al., 2015). Dit geldt niet alleen voor vrouwen en/of mensen van kleur, maar ook voor bijvoorbeeld ouderen die bang zijn om aan een stereotiep beeld te voldoen (Barber, 2020).

Een verklaring voor *stereotype threat* en verminderde presentatie, ligt volgens een review van Van Landingham en anderen (2022) in het feit dat de angst om te voldoen aan negatieve stereotypen, er toe kan leiden dat cognitief dan de focus ligt op om hier niet aan te voldoen, waardoor de ruimte kan afnemen voor andere cognitieve processen, die nodig zijn om bepaalde taken goed uit te voeren. Simpel gezegd: als je bang bent om aan een stereotiep te voldoen, dan ben je daar mee bezig, dus enigszins afgeleid waardoor je minder goed kan presteren.

Werkgevers kunnen als oplossing proberen ervoor te zorgen dat deelnemers meer op hun gemak zijn tijdens de test, zo wordt geconcludeerd in een review van Walton en collega's (2015). Tegelijkertijd wordt gesteld dat een test die bepaalt hoe intelligent je bent per definitie een situatie kan veroorzaken dat mensen bang zijn om stereotype beelden te bevestigen over zichzelf. Het minder laten meewegen van zo'n test – of zelfs geheel niet afnemen van een test – lijkt dus meer voor de hand te liggen (Walton et al., 2015).

Stereotype threat is ook een van de mogelijke verklaringen voor Zwarte Amerikanen minder goed presteren op IQ-testen, zo blijkt uit een studie van McKay (et al., 2002) en in een studie van McKay (et al., 2003). Uit een studie van Brown en Day (2006) blijkt zelfs dat wanneer Zwarte Amerikanen zich niet bedreigd voelen (de test is aangekondigd als gewoon een puzzel en er wordt niet gesproken over het meten van intelligentie), dat zij hetzelfde scoren als witte Amerikanen.

## 2.5 Dragen intelligentietesten bij aan verschillende kansen in selectieprocessen?

Uit het voorgaande werd duidelijk dat intelligentietesten verschillende resultaten kunnen opleveren voor mensen met verschillende culturele achtergronden, met verschillende beperkingen en genders. De vraag is vervolgens of deze verschillen er ook aan bijdragen dat mensen vanwege deze kenmerken minder kans maken op selectie bij een sollicitatieprocedure. En zijn als decennia studies die vaak laten zien dat dit inderdaad het geval is (bijvoorbeeld LeBlanc et al., 2001). Een recente meta-analyse uit het Verenigd Koninkrijk toont dit ook aan: het bleek dat verschillen op intelligentietesten bij droegen aan lagere selectiepercentages bij sollicitanten met een migratieachtergrond. In het Verenigd Koninkrijk bleken intelligentietesten dus te kunnen bijdragen aan ongelijke kansen op de arbeidsmarkt (Te Nijenhuis et al., 2024).

### Samenvatting hoofdstuk 2

Het gebruik van gestandaardiseerde intelligentietesten (IQ-testen of capaciteitentesten) is gemeengoed, en studies laten zien dat dergelijke testen valide gebruikt zouden kunnen worden voor respondenten met verschillende culturele achtergronden. Desalniettemin blijkt dat mensen met een migratieachtergrond gemiddeld gezien lager scoren op deze testen. Niet helemaal duidelijk is waarom. Een mogelijke verklaring is dat veel intelligentietesten heel talig zijn in hun opzet. Dit kan problemen opleveren voor mensen waarbij de taal van de test niet hun moedertaal is. Ook het aantal jaar in het onderwijssysteem van het betreffende land kan een rol spelen;

mogelijk leert men daarbij niet alleen de taal, maar ook de culturele referenties die in een test verweven zitten.

Ook voor kinderen met dyslexie is de taligheid van de intelligentietesten een obstakel. Daarnaast is het voor een goede score op deze testen ook belangrijk om je goed te kunnen concentreren. Dat kan lastiger zijn voor mensen met ADHD. Maar focus kan ook gehinderd worden door stereotype threat: door bang te zijn om aan een stereotiep beeld te voldoen als minder intelligent, kunnen mensen die behoren tot deze gestereotypeerde groepen (zoals mensen van kleur, ouderen, maar ook bijvoorbeeld vrouwen als het gaat om wiskunde) minder goed presteren op IQ-testen. Deze 'stereotype threat' is in V.S. aangetoond als een plausibele verklaring voor het verschil in prestaties op een IQ-test tussen witte en Zwarte Amerikanen.

Verschiedende groepen mensen kunnen dus lager scoren op een intelligentietest vanwege hun achtergrond. En omdat deze testen ook vaak een belangrijke rol spelen in de selectie van kandidaten heeft dit ook negatieve effecten voor de baankansen van deze groepen. Bovendien blijkt uit recent onderzoek dat intelligentietesten veel minder voorspellend zijn voor prestaties op het werk dan eerder werd gedacht. Andere methoden, zoals gestructureerd interviewen en het toetsen van vakkennis, zijn betere voorspellers van toekomstige prestaties, daarbij zijn intelligentietoetsen zelfs niet meer van toegevoegde waarde.

# Hoofdstuk 3: Casuïstiek

We zijn op zoek gegaan naar casussen die gaan over assessments en gelijk behandeling en inclusie. Hiervoor hebben we gezocht in de database van het College voor de Rechten van de Mens en gezocht op de trefwoorden ‘assessment’ en ‘test’.

## 3.1 De gevonden casussen

Er zijn vier casussen gevonden die we hieronder beschrijven. De volledige beschrijving is te vinden op de site van het College voor de Rechten van de Mens. Hierbij moet wel de kanttekening worden gemaakt dat oordelen van het College voor de Rechten van de Mens alleen de juridische definitie van ‘onderscheid’ betreffen. De ervaren en de daadwerkelijke discriminatie wordt hiermee niet in beeld gebracht. Bovendien betreffen de gepubliceerde oordelen geen onderscheid tussen mannen en vrouwen. Bij casus 2 en casus 3 gaat het om dezelfde sollicitant, waarbij zowel het bureau voor arbeidsbemiddeling (casus 2) als de potentiële werkgever (casus 3) zijn onderzocht.

**Casus 1: Geen discriminatie door Koninklijke Vezet B.V. van een man met autisme door hem uit een leidinggevende functie te laten plaatsen (zie [deze link](#) voor de hele casus)**

- Het Koninklijke Vezet heeft een leidinggevende uit zijn functie geplaatst en zeggen hierbij dat ze uit zijn competentieprofiel na een assessment opmaken dat zijn zwakste competenties motiveren, overtuigen, teambuilding en omgevings sensitiviteit zijn.
- Volgens de man is dat omdat hij ASS (autismespectrumstoornis) heeft, Daarnaast voert de man aan dat het gebruik van het competentieprofiel een onderscheidmakende methode is.
- Het College voor de Rechten van de Mens oordeelt dat een medewerker op zichzelf geschikt moet zijn voor de uitoefening van de functie, hierdoor hoeft een werkgever geen rekening te houden met zijn beperking bij het vaststellen van zijn geschiktheid.
- Daarnaast oordeelt het College dat Koninklijke Vezet B.V. niet is tekortgeschoten in het aanbieden van doeltreffende aanpassingen. Deze aanpassingen bleken niet voldoende om de belemmeringen weg te nemen om de functie uit te kunnen voeren.

**Casus 2: Meurs HRM B.V. heeft verboden onderscheid op grond van handicap of chronische ziekte gemaakt bij de arbeidsbemiddeling bij de afwijzing van een sollicitant door niet te onderzoeken of de sollicitant behoefte had aan een doeltreffende aanpassing en zo ja, welke (zie [deze link voor de hele casus](#))**

- Een kandidate moest een capaciteitentest en een persoonlijkheidstest doen bij de selectie voor een traineeprogramma.
- De vrouw had een beperking die invloed heeft op ruimtelijk en abstract inzicht. Dat heeft zij zoals gevraagd ook aangegeven.
- Ze scoorde te laag en werd afgewezen op basis van de lage score op de testen, die door het arbeidsbemiddelingsbureau werden afgenomen en gescoord.
- Oordeel is dat bij de beoordeling van haar testresultaten geen onderzoek is gedaan naar de behoefte van doeltreffende aanpassingen naar aanleiding van het melden van haar beperking en haar lage score. Er is daarom geen rekening gehouden met haar beperking en dus is dit een verboden onderscheid.

**Casus 3: De gemeente Rotterdam heeft verboden onderscheid op grond van handicap of chronische ziekte gemaakt bij de selectie bij de afwijzing van een sollicitant door niet te onderzoeken of de sollicitant behoefte had aan een doeltreffende aanpassing en zo ja, welke (zie [deze link voor deze hele casus](#))**

- Een vrouw heeft gesolliciteerd naar het traineeprogramma voor beleidsmedewerker bij de gemeente Rotterdam. De gemeente heeft de selectie voor een groot deel overgedragen aan een selectiebureau.

De kandidaten moesten een aantal testen maken, waaronder een capaciteitentest en een persoonlijkheidstest.

- De vrouw heeft een *Non Verbal Learning Disorder* (NLD). Hierdoor heeft zij beperkingen bij ruimtelijk en abstract inzicht. Zij kan daarom testen waarin deze inzichten worden getoetst minder goed maken dan anderen.
- In de begeleidende informatie bij de testen was vermeld dat kandidaten die een beperking hebben die van invloed kan zijn op de resultaten, hiervan melding moesten maken. Het selectiebureau kon hiermee dan rekening houden bij de beoordeling van de testresultaten. Nadat de vrouw de capaciteitentest had gemaakt, heeft zij het selectiebureau een mail gestuurd dat zij een NLD heeft.
- Daarna heeft het selectiebureau de vrouw meegedeeld dat zij te laag had gescoord op de capaciteitentest en dat zij daarom was afgewezen.
- Het College voor de Rechten van de Mens spreekt als zijn oordeel uit dat de gemeente Rotterdam jegens de vrouw verboden onderscheid heeft gemaakt op grond van handicap of chronische ziekte bij de afwijzing voor het traineeprogramma. De gemeente Rotterdam is volgens het oordeel eindverantwoordelijk voor het handelen van het selectiebureau (zie casus 2). Omdat het selectiebureau geen rekening heeft gehouden met de beperking van de vrouw in het beoordelen van haar testresultaten, is ook de gemeente verantwoordelijk voor het verboden onderscheid op basis van deze beperking.

**Casus 4: Niet gebleken dat een wervings- en selectiebureau onderscheid op grond van ras en/of leeftijd heeft gemaakt door haar betrokkenheid bij de afwijzing van een man voor de functie van Rijkstrainee (zie [deze link voor de hele casus](#))**

- Een man heeft gesolliciteerd naar de functie van Rijkstrainee. Bij het wervings- en selectiebureau heeft hij een cognitieve capaciteitentest en een persoonlijkheidstest afgelegd. Mede op basis van de testresultaten heeft een ministerie de man vervolgens uitgenodigd voor een selectie-interview.
- In de rapportage van dit interview worden de competenties van de man voor de functie van Rijkstrainee overwegend als matig beoordeeld. Op grond hiervan heeft het ministerie de man afgewezen.
- Het College voor de Rechten van de Mens heeft geen oordeel geveld over de betrokkenheid van het bureau bij de afwijzing van de kandidaat. Het College stelt dat er geen feiten zijn aangevoerd waaruit blijkt dat het wervings- en selectiebureau onderscheid op grond van ras en/of leeftijd heeft gemaakt bij de arbeidsbemiddeling. Het onderscheid komt hier niet mee vast te staan.

### 3.2 Wat de casussen laten zien

Uit deze casussen komt naar voren dat kandidaten afgewezen kunnen worden voor een specifieke baan op basis van een assessment. In de zaken dat het College voor de Rechten van de Mens oordeelt dat er verboden onderscheid op grond van handicap en chronische ziekte is gemaakt, gaat het om een assessment met een persoonlijkheidstest en een intelligentie-

test/capaciteitentest (casus 2 en casus 3). De betreffende kandidaat heeft vooraf aangegeven een beperking te hebben, maar er werd niet onderzocht op welke manier er aanpassingen konden worden gedaan om in de test of de beoordeling daarvan rekening te houden met de beperking. De lage scores van de kandidaat werden vervolgens gebruikt in de beoordeling en afwijzing voor de functie. En dat is dus in strijd met de gelijk behandelingswetgeving.

Als echter een kandidaat wordt afgewezen in een assessment op basis van het niet voldoen aan competenties die noodzakelijk zijn voor essentiële taken van de functie, en waarbij deze belemmering niet kan worden weggenomen door doeltreffende aanpassing, dan wordt er geen onterecht onderscheid gemaakt en is er dus geen sprake van discriminatie (casus 1), aldus het College voor de Rechten van de Mens. En wanneer iemand wordt afgewezen op basis van een assessment, maar niet kan onderbouwen waarom dit een vorm van discriminatie is, kan niet worden vastgesteld of er sprake is van ongerechtvaardigd onderscheid, volgens het College voor de Rechten van de Mens (casus 4).

Kortom, de huidige situatie is dus zo dat een assessment wordt beschouwd als een vorm van discriminatie wanneer concreet kan worden aangetoond hoe die betreffende assessment heeft geleid tot een verboden onderscheid op een specifiek te duiden discriminatiegrond.

# Conclusies en aanbevelingen

In dit verkennende literatuuronderzoek zijn we nagegaan of er mogelijke risico's zijn aan het gebruik van assessments voor selectie en promotie vanuit het oogpunt van gelijke behandeling en inclusie. In andere woorden hebben we onderzocht of, wanneer je een inclusieve organisatie bent, het gebruik van assessments verstandig is.

Onze hoofdvraag: Bestaan er risico's dat het gebruik van assessments in selectie- en promotieprocedures leidt tot ongelijke behandeling en uitsluiting?

## Deelvragen

1. Zijn er risico's als het gaat om gelijke behandeling of inclusie, bij het gebruik van psychologische testen in assessments?
2. Zijn er risico's als het gaat om gelijke behandeling of inclusie bij het gebruik van IQ-testen in assessments?
3. Zijn er casussen bekend uit de praktijk waarin er risico's naar voren zijn gekomen als het gaat om gelijke behandeling en inclusie met betrekking tot assessments voor selectie en promotie doeleinden?

Hieronder beantwoorden we eerst kort de hoofdvraag en daarna de deelvragen.

## Hoofdvraag kort beantwoord

In een notendop kan als antwoord op de hoofdvraag worden gesteld dat op basis van deze verkenning van de literatuur, voorzichtigheid is geboden bij het gebruik van psychologische testen en intelligentietesten als onderdeel van assessments vanuit het oogpunt van inclusie en gelijke behandeling.

- Wanneer deze testen worden ingezet in assessments door bedrijven die juist streven naar inclusie is er het risico dat dit juist leidt tot minder inclusie, omdat bepaalde groepen mensen gemiddeld minder scoren op deze testen. Het gaat om mensen die opgegroeid zijn in een niet-Nederlands schoolsysteem, de Nederlandse taal niet als moedertaal hebben, meer schakelen tussen verschillende culturele contexten én behoren tot een gestereotypeerde groep. Dit kan minder betrouwbare scores opleveren op hun psychologische test, onder meer omdat in verschillende culturele contexten er verschillende opvattingen kunnen zijn over persoonlijkheid en persoonlijkheid niet vaststaand is en vooral mensen uit meerdere culturen hierin kunnen schakelen. Dit kan erin resulteren dat mensen die zich in meerdere sociale contexten bewegen en/of binnen een andere sociale context zijn opgegroeid dan



de Westerse, onterecht lager scoren op hun intelligentietesten. In de praktijk kan betekenen dat mensen minder kans maken om door een assessment komen, mede vanwege hun culturele achtergrond.

- Wanneer deze testen worden gebruikt om te bepalen of iemand wel of niet de baan krijgt, dan maken onder meer kandidaten met een migratieachtergrond gemiddeld gezien dus minder kans.
- Dat geldt mogelijk ook voor ouderen, vrouwen en andere groepen waarover stereotype beelden kunnen leven in deze samenleving, alsof zij minder intelligent zouden zijn: want als je bang bent te voldoen aan een stereotype, kan die angst ervoor zorgen dat je het minder goed doet op een intelligentietest.
- Het grootste knelpunt spelen op het gebied van neurodiversiteit (verschillen in hoe je brein werkt). Zo kunnen mensen met dyslexie lager scoren op IQ-testen. En de score van mensen met bijvoorbeeld ADHD of autisme kan als minder positief worden gezien op een psychologisch test. Bovendien kan het werken met een psychologische test discriminatie ten aanzien van hen in de hand werken, doordat deze testen hun neurodivergentie onthullen aan de werkgever.
- Er kunnen daarnaast vraagtekens worden geplaatst bij de voorspellende waarde van persoonlijkheids- en intelligentietesten. Recente studies laten zien dat gestructureerd interviewen en toetsen op werkgerelateerde vaardigheden of vakkennis een veel betere voorspelling geven van iemands prestaties op het werk en dat persoonlijkheids- en intelligentietesten daar weinig toegevoegde waarde bij hebben.
- Uit de casussen van het College voor de Rechten van de Mens blijkt dat assessments in de praktijk gebruikt worden in afwijzingen

van kandidaten en dat wanneer er niet voldoende rekening wordt gehouden met een beperking in de testen of de behaalde resultaten, dat dit leidt tot een verboden onderscheid op een specifiek te duiden discriminatiegrond.

### **Psychologische testen en risico's met betrekking tot gelijke behandeling en inclusie (antwoord deelvraag 1)**

Persoonlijkheidstesten kunnen in enige mate voorspellend zijn voor de prestatie van iemand in een bepaalde baan. Maar hoe sterk dat het geval is, hangt sterk af van het type werk en de context. Desalniettemin lijken deze testen wel gebruikt te worden met het doeleinde het presteren van iemand te voorspellen in een bepaalde baan.

Er zijn enige risico's verbonden aan het gebruik van psychologische testen als het gaat om inclusie en gelijke behandeling. Als het gaat om gender, komt uit de literatuur naar voren dat het kan zijn dat vrouwen en mannen anders scoren op deze testen, maar de literatuur is er niet eenduidig over. Hetzelfde geldt als het gaat om gelijke behandeling op grond van (culturele) achtergrond. Er is enerzijds bewijs dat de veel gebruikte testen, gebaseerd op de psychologische theorie van de The Big Five kenmerken, in verschillende culturele contexten bruikbaar zijn, maar er is ook bewijs van het tegendeel. Zo is aangetoond dat bepaalde eigenschappen anders gewaardeerd worden in verschillende culturele contexten en wordt de Big Five daarom niet overal als compleet wordt ervaren.

Daarnaast is er meer fundamentele kritiek dat persoonlijkheid wordt gezien als iets vaststaands en onveranderbaars, terwijl er aanwijzingen zijn dat persoonlijkheid bijvoorbeeld verschilt per context, wat mogelijk dus kan betekenen dat bijvoorbeeld een test afnemen thuis tot andere resultaten leidt dan afnemen op het werk. Daarmee is er mogelijk het risico dat een persoonlijkheidstest een verkeerd beeld geeft van iemand die meer het eigen gedrag aanpast in verschillende contexten. Dit geldt mogelijk in het bijzonder voor mensen die zijn opgegroeid zijn met verschillende culturele achtergronden gewend zijn om hiertussen te schakelen. Bovendien is de waarde van zo'n psychologische test mogelijk ook minder omdat deze test vervalst kan worden door de antwoorden zo in te vullen dat deze passen bij wat er wordt gezocht. En de mate waarin mensen met verschillende culturele achtergronden hun antwoorden aanpassen, kan verschillen. Het is moeilijk deze sociale wenselijkheid te controleren. En daarbij is het ook de vraag of een dergelijk aanpassingsvermogen misschien juist wenselijk is voor bepaalde banen.

De grootste kritiek als het gaat om persoonlijkheidstesten – in het bijzonder de veel gebruikte testen gebaseerd op de Big Five – is dat deze mensen met een beperking kunnen benadelen. Het gaat dan in het bijzonder om mensen die kunnen worden beschouwd als neurodivergent. Een werkgever die bijvoorbeeld niet de baan wil geven aan iemand die autistisch is, ADHD heeft of een angststoornis, kan op basis van de uitslagen van deze test deze mensen herkennen als dusdanig. Op die manier kan een psychologische test de discriminatie van neurodivergente mensen in de hand werken. Maar ook als dat niet gebeurt, kunnen neurodivergente mensen

minder goed uit zo'n test komen: er komt bijvoorbeeld bij iemand met een angststoornis uit dat deze erg angstig is en dat kan door de werkgever worden gezien als minder wenselijk.

### **Intelligentietesten en risico's met betrekking tot gelijke behandeling en inclusie (antwoord deelvraag 2)**

Intelligentietesten worden gezien als een belangrijke voorspeller van iemands prestaties op het werk. Intelligentietesten blijken uit recent onderzoek inderdaad werkprestaties te voorspellen. Maar gestructureerd interviewen en testen die meer ingaan op vaardigheden of kennis die direct verband houden met het werk blijken een veel belangrijkere voorspeller, zelfs in die mate dat intelligentietesten geen toegevoegde waarde hebben.

Hoewel gestandaardiseerde intelligentietesten doorgaans een valide manier zijn om iets te kunnen zeggen over intelligentie, kunnen dit soort testen toch een obstakel vormen voor gelijke behandeling op de arbeidsmarkt. Onderwijs speelt een belangrijke rol in hoe goed mensen zijn in het maken van dergelijke testen. Niet alleen kan onderwijs mensen de taal van de toets aanleren, maar hen wellicht ook gewend maken aan bepaalde culturele referenties of het type vragen dat wordt gesteld. Vandaar dat voor migranten het aantal jaar onderwijs in het land waar men de test maakt ook een belangrijke voorspeller is van hun scores op dergelijken toetsen. Naar onderwijs speelt ook taal een rol. Voor veel intelligentietesten is een goed taalbegrip belangrijk voor het begrijpen van de vragen. Dit kan een risico op lagere scores vormen voor migranten voor wie dit

niet hun moedertaal is, maar ook voor bijvoorbeeld dyslectici. En net als bij persoonlijkheidstesten, blijken intelligentietesten potentieel nadelen voor neurodivergente personen. Zo kan dyscalculie een obstakel vormen bij rekenopgaven, en kan ADHD een obstakel zijn voor de concentratie die nodig is om goed te presteren op intelligentietesten.

Een specifiek probleem voor intelligentietesten is dat het bewustzijn van stereotype beelden over de groep waar jezelf bij hoort, kunnen hinderen in het maken van dit soort testen. Wanneer mensen, zelfs op heel subtiele manieren (zoals het invullen van gender of culturele achtergrond op een formulier of de aankondiging dat de test hun intelligentie meet), herinnerd worden aan het stereotiepe beeld over hun groep (denk aan stereotiepe beelden over dat de vrouwen minder goed zouden zijn in wiskunde, dat ouderen niet snel genoeg zouden denken etc.), kunnen zij daardoor angst ervaren om het stereotiepe beeld te bevestigen. Deze angst kan een hinderende factor zijn bij het maken van de toets en zorgen voor een lagere score. Deze angst is dus ook een van de verklaringen dat bepaalde groepen gemiddeld lager scoren op IQ-testen.

### **Casussen bekend uit de praktijk en risico's met betrekking tot gelijke behandeling en inclusie (antwoord deelvraag 3)**

We hebben vier casussen gevonden in de database van het College voor de Rechten van de Mens met oordelen, door te zoeken op de zoektermen 'assessments' en 'test', die gaan over oordelen ten aanzien van situaties waarin een kandidaat aangeeft te zijn gediscrimineerd in een assessment situatie. Op basis daarvan kan allereerst geconcludeerd worden dat het in

de praktijk zo kan zijn dat de score op een psychologische test en op een intelligentietest (mede)bepalend is voor of iemand een bepaalde baan wel of niet krijgt. Ten tweede kan geconcludeerd worden dat het wordt gezien als onterecht onderscheid (ofwel discriminatie) wanneer kandidaten een test moeten doen in het kader van hun assessment, vooraf hebben aangegeven een beperking te hebben die ervoor kan zorgen dat de resultaten minder positief zijn, en waarbij er niet wordt onderzocht op welke manier er aanpassingen kunnen worden gedaan om rekening te houden met deze beperking. De afwijzing die wordt gemaakt op basis van test scores die geen rekening houden met de beperking, wordt beoordeeld als onterecht onderscheid. Als echter een kandidaat wordt afgewezen in een assessment op basis van het niet voldoen aan noodzakelijk competenties voor essentiële taken binnen de functie (en er geen doeltreffende aanpassingen mogelijk zijn), dan wordt er geen onterecht onderscheid gemaakt en is er dus geen sprake van discriminatie. Dat betekent dat de huidige situatie zo is dat een assessment wordt beschouwd als een vorm van discriminatie wanneer concreet kan worden aangetoond hoe die betreffende assessment heeft geleid tot een verboden onderscheid op een specifiek te duiden discriminatiegrond.

### **Aanbevelingen**

Onze belangrijkste aanbeveling voor werkgevers die inclusie nastreven, is om het gebruik van assessments kritisch te heroverwegen en de risico's in ogenschouw te nemen. Een systematische review en/of meta-analyse als opvolging van deze eerste verkenning is zeer aan te raden om preciezer

zicht te krijgen op de verschillende risico's met betrekking tot inclusie en gelijke behandeling (en de grootte daarvan) bij het gebruik van zowel intelligentietesten als psychologische testen voor assessment doeleinden. Tevens is aan te raden om onderzoek te doen naar de onderdelen van assessments die niet zijn meegenomen in deze verkenning, zoals rollenspellen en praktijkopdrachten.

Ook is aan te raden om preciezer na te gaan hoe de methode van objectief werven en selecteren, wel of niet gecombineerd kan worden met assessments. Bij objectief werven en selecteren staan vier tot zes competenties centraal voor de betreffende functie en wordt een kandidaat enkel en alleen daarop afgerekend (VIA, 2023). De vraag is of het afnemen van een psychologische test en een intelligentietest, die meer een totaalbeeld schetsen van de psychologische kenmerken en intelligentie van een kandidaat, hier niet haaks op staan. Verder onderzoek zou dit moeten uitwijzen. Onderzoek lijkt nu te suggereren dat assessments van algemene niet werk-gerelateerde kenmerken weinig bijdraagt aan het voorspellen van de toekomstige werkprestaties van een kandidaat. Ook is onduidelijk uit de literatuur of het moment waarop deze assessments worden ingezet van invloed is op de invloed die ze hebben in het selectieproces. Dat vereist ook meer onderzoek.

In dat kader is het ook aan te bevelen om actief op zoek te gaan naar alternatieve objectievere vormen van beoordeling, zoals concrete opdrachten die specifiek passen bij de betreffende functie. Onderzoek naar hoe dat precies het beste uit te voeren, is aan te bevelen.

Op basis van deze verkenning van de literatuur blijven daarnaast ook nog juridische vragen onbeantwoord. Nader onderzoek zou nodig zijn om na te gaan of en wanneer assessments precies wel of niet voldoen aan de wetgeving met betrekking tot gelijke behandeling. In hoeverre is het kenbaar moeten maken van een beperking voorafgaand aan een assessment en tevens het onthuld worden van een beperking (of neurodivergentie) door het doen van een assessment wel of niet in strijd met de gelijke behandelingswetgeving en privacywetgeving?

Een daaraan gerelateerde vraag is: hoe om te gaan met beperkingen die niet als dusdanig gediagnosticeerd zijn? Zo wordt nu bij assessments aan kandidaten gevraagd om vooraf kenbaar te maken dat zij een beperking hebben, maar onduidelijk is bijvoorbeeld hoe om te gaan met het feit dat niet iedereen een precieze diagnose heeft voor neurodivergentie: zo is bijvoorbeeld autisme bij vrouwen vaak niet gediagnosticeerd, zo meldt de Nederlandse Vereniging voor Autisme (z.d.).

# Bronnen

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of cross-cultural psychology*, 35(1), 13-28.

Allik J, Realo A and McCrae RR (2023). Conceptual and methodological issues in the study of the personality-and-culture relationship. *Front. Psychol.* 14:1077851. doi: 10.3389/fpsyg.2023.1077851.

Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: Review and meta-analysis. *Frontiers in Psychology*, 6, 900.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., ... & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of personality and social psychology*, 86(2), 356.

Barber, S. J. (2020). The applied implications of age-based stereotype threat for older adults. *Journal of Applied Research in Memory and Cognition*, 9(3), 274-285.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1), 1-26.

Bell, L. J., John, O. P., & Hinshaw, S. P. (2024). ADHD Symptoms in Childhood and Big Five Personality Traits in Adolescence: A Five-Year Longitudinal Study in Girls. *Research on Child and Adolescent Psychopathology*, 1-14.

Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of personality and social psychology*, 75(3), 729.

Berry, C. M., Lievens, F., Zhang, C., & Sackett, P. R. (2024). Insights from an updated personnel selection meta-analytic matrix: Revisiting general mental ability tests' role in the validity-diversity trade-off. *Journal of Applied Psychology*.

Bloemers, W. (2013). *De nieuwe assessmentgids* Amsterdam: Uitgeverij Ambo.

Bridgett, D. J., & Walker, M. E. (2006). Intellectual functioning in adults with ADHD: a meta-analytic examination of full scale IQ differences between adults with and without ADHD. *Psychological assessment*, 18(1), 1.

- Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: stereotype threat and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology, 91*(4), 979.
- Brown, T. E., Reichel, P. C., & Quinlan, D. M. (2009). Executive function impairments in high IQ adults with ADHD. *Journal of Attention Disorders, 13*(2), 161-167.
- Chen, S. X., Benet-Martínez, V., & Harris Bond, M. (2008). Bicultural Identity, bilingualism, and psychological adjustment in multicultural societies: immigration-based and globalization-based acculturation. *Journal of personality, 76*(4), 803-838.
- Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin, 36*(11), 1514-1528.
- Cheung, F. M., Leung, K., Zhang, J. X., Sun, H. F., Gan, Y. Q., Song, W. Z., & Xie, D. (2001). Indigenous Chinese personality constructs: Is the five-factor model complete?. *Journal of cross-cultural psychology, 32*(4), 407-433.
- Church, A. T., Willmore, S. L., Anderson, A. T., Ochiai, M., Porter, N., Mateo, N. J., ... & Ortiz, F. A. (2012). Cultural differences in implicit theories and self-perceptions of traitedness: Replication and extension with alternative measurement formats and cultural dimensions. *Journal of Cross-Cultural Psychology, 43*(8), 1268-1296.
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment, 4*(1), 5-13.
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics. University of Chicago Legal Forum, 139.
- Day, M. & Badou, M. (2019). *Geboren en getogen. Onderzoek naar de identiteitsbeleving en gevoelens van binding van jongeren met een migratie-achtergrond*. Utrecht: Kennisplatform Inclusief Samenleven.
- De Kock, X., Desoete, A., & Stock, P. (2008). De relatie tussen intelligentie en rekenvaardigheden. *Scriptie neergelegd tot het behalen van de graad van Licentiaat in de Psychologie, optie Klinische Psychologie*. Gent: Universiteit van Gent.
- Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance, 16*(1), 81-106.
- Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological science, 16*(12), 939-944.

Eerkens, M. (2015) *Domme vraag, maar wat is IQ eigenlijk?* Geraadpleegd op 1-11-2024 van <https://decorrespondent.nl/3365/domme-vraag-maar-wat-is-iq-eigenlijk/bdf38a3b-4e2e-0b5f-3cb4-c47fed7bebb1>

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*(2), 155-166.

Felten, H. & Broekroelofs, R. (2022) *Wat werkt bij het verminderen van discriminatie?* Utrecht: Kennisplatform Inclusief Samenleven & Movisie.

Felten, H. (2022). Wat is objectief werven en selecteren? Geraadpleegd op 3-10-2024 van <https://www.kis.nl/artikel/wat-objectief-werven-en-selecteren>.

Froehlich, L., Mok, S. Y., Martiny, S. E., & Deaux, K. (2022). Stereotype threat-effects for Turkish-origin migrants in Germany: Taking stock of cumulative research evidence. *European Educational Research Journal, 21*(2), 330-354.

Gage, E., Coker, S., & Jobson, L. (2015). Cross-cultural differences in desirable and undesirable forms of self-consistency and influence on symptoms of depression and anxiety. *Journal of Cross-Cultural Psychology, 46*(5), 713-722.

Gordijn, E. & Wigboldus, D. (2013) Stereotypen. In: *Sociale psychologie*. Groningen / Houten: Noordhoff Uitgevers.

Griffith, R. L., Chmielowski, T., Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*(3), 341-355.

Gurven, M., Von Rueden, C., Massenkoff, M., Kaplan, H., & Lero Vie, M. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager-farmers in the Bolivian Amazon. *Journal of personality and social psychology, 104*(2), 354.

Haft, S. L., Greiner de Magalhães, C., & Hoeft, F. (2023). A systematic review of the consequences of stigma and stereotype threat for individuals with specific learning disabilities. *Journal of Learning Disabilities, 56*(3), 193-209.

Harkness, A. R., McNulty, J. L., & Ben-Porath, Y. S. (1995). The personality psychopathology five (PSY-5): Constructs and MMPI-2 scales. *Psychological assessment, 7*(1), 104.

Hausdorf, P. A., & Robie, C. (2018). Understanding subgroup differences with general mental ability tests in employment selection: Exploring socio-cultural factors across inter-generational groups. *International Journal of Selection and Assessment, 26*(2-4), 176-190.

Heaven, P. C., & Ciarrochi, J. (2012). When IQ is not everything: Intelligence, personality and academic performance at school. *Personality and Individual Differences, 53*(4), 518-522.

Holman, D. J., and Hughes, D. J. (2021). Transactions between Big-5 personality traits and job characteristics across 20 years. *J. Occup. Organ. Psychol.* 94, 762-788. doi: 10.1111/joop.12332.

Hulsegge, G., Emmert, S., & Verhoef, H. (2020). *Delphi-studie Stand der Wetenschap Arbeidsmarktdiscriminatie: Onderzoek naar de belangrijkste werkzame interventies om arbeidsmarktdiscriminatie te verminderen en de diversiteit te vergroten tijdens werving en selectie*. Leiden: TNO.

Jansen, T. (2019). *Leiders selecteren op basis van IQ-testen? Dat is misschien niet zo zinvol. Leiders selecteren op basis IQ-testen? Dat is misschien niet zo zinvol* | MT/Sprout via mtsprout.nl.

Johnston, A., & Kent, J. L. (2019). The relationship between leadership strengths and job performance of fleet managers. *Transportation journal*, 58(2), 126-144.

Judge, T. A., & Zapata, C. P. (2015). The person–situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal*, 58(4), 1149-1179.

Kemmelmeier, M., & Kühnen, U. (2012). Culture as process: The dynamics of cultural stability and change. *Social Psychology*.

Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology*, 93(1), 140-154.

Kovacs, K., & Conway, A. R. (2019). What is IQ? Life beyond “general intelligence”. *Current Directions in Psychological Science*, 28(2), 189-194.

Kubiak, E., Efremova, M. I., Baron, S., & Frasca, K. J. (2023). Gender equity in hiring: examining the effectiveness of a personality-based algorithm. *Frontiers in Psychology*, 14, 1219865.

Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., ... & Vakis, R. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. *Science advances*, 5(7), eaaw5226.

Lodi-Smith, J., Rodgers, J. D., Cunningham, S. A., Lopata, C., & Thomeer, M. L. (2019). Meta-analysis of Big Five personality traits in autism spectrum disorder. *Autism*, 23(3), 556-565.

Lönnqvist, J. E., Konstabel, K., Lönnqvist, N., and Verkasalo, M. (2014). Accuracy, consensus, in-group bias, and cultural frame shifting in the context of national character stereotypes. *J. Soc. Psychol.* 154, 40-58. doi: 10.1080/00224545.2013.843500.



McCrae, R. R., & Terracciano, A. (2005). Personality profiles of cultures: aggregate personality traits. *Journal of personality and social psychology*, 89(3), 407.

McGee, A., & McGee, P. (2022). *Gender and race differences on incentivized personality measures*. Working paper.

McKay, P. F., Doverspike, D., Bowen-Hilton, D., & Martin, Q. D. (2002). Stereotype threat effects on the raven advanced progressive matrices scores of African Americans 1. *Journal of Applied Social Psychology*, 32(4), 767-787.

McKay, P. F., Doverspike, D., Bowen-Hilton, D., & McKay, Q. D. (2003). The effects of demographic variables and stereotype threat on Black/White differences in cognitive ability test performance. *Journal of Business and Psychology*, 18, 1-14.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel psychology*, 60(3), 683-729.

Nationale Vacaturebank (z.d.). *Het assessment oefenen en voorbereiden* Geraadpleegd op 30-1-2024 van <https://www.nationalevacaturebank.nl/carriere/sollicitatiegesprek/kennismakingsgesprek/het-assessment-oefenen-en-voorbereiden?referrer=https%3A%2F%2Fwww.google.nl%2F>.

Nettelbeck, T., & Wilson, C. (2005). Intelligence and IQ: What teachers should know. *Educational Psychology*, 25(6), 609-630.

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of applied psychology*, 93(6), 1314.

Nigg, J. T., John, O. P., Blaskey, L. G., Huang-Pollock, C. L., Willcutt, E. G., Hinshaw, S. P., & Pennington, B. (2002). Big five dimensions and ADHD symptoms: links between personality traits and clinical symptoms. *Journal of personality and social psychology*, 83(2), 451.

Nederlandse Vereniging voor Autisme (z.d.). *Vrouwen*. Geraadpleegd op 30-10-2024 van <https://www.autisme.nl/over-autisme/doelgroepen/vrouwen/#:~:text=Autisme%20bij%20vrouwen%2C%20zeker%20ook,symptomen%20van%20autisme%20beter%20camoufleren>.

Pesta, B. J., Kirkegaard, E. O., te Nijenhuis, J., Lasker, J., & Fuerst, J. G. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence*, 78, 101408.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61(1), 153–172. <https://doi.org/10.1111/j.1744-6570.2008.00109.x>.

- Powell, D. M., Goffin, R. D., & Gellatly, I. R. (2011). Gender differences in personality scores: Implications for differential hiring rates. *Personality and Individual Differences, 50*(1), 106-110.
- Randstad (z..d.). Assessment: zo bereid jij je voor. Geraadpleegd op 29-8-2024 van <https://www.randstad.nl/ontwikkelen/loopbaan/solliciteren/assessment>.
- Razani, J., Murcia, G., Tabares, J., & Wong, J. (2007). The effects of culture on WASI test performance in ethnically diverse individuals. *The Clinical Neuropsychologist, 21*(5), 776-788.
- Remann, P., & Nordin, A. (2021). Personality tests in recruitment. *Ed. project, 2*-15.
- Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). The problem of bias in psychological assessment. In *Mastering Modern Psychological Testing: Theory and Methods* (pp. 573-613). Cham: Springer International Publishing.
- Robinson, E., Hull, L., & Petrides, K. V. (2020). Big Five model and trait emotional intelligence in camouflaging behaviours in autism. *Personality and Individual Differences, 152*, 109565.
- Ronner, S., & Noks Nauta, A. P. (2010). Een talent voor valkuilen: De psychologie van hoogbegaafdheid. *Psychopraktijk, 2*, 13-16.
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology, 107*(11), 2040.
- Schmitt, D. P., Allik, J., McCrae, R. R., Benet-Martinez, V., Alcalay, L., Ault, L., ... Zupancic, A. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology, 38*(2), 173-212.
- Schmitt, D. P., Long, A. E., McPhearson, A., O'Brien, K., Remmert, B., & Shah, S. H. (2017). Personality and gender differences in global perspective. *International Journal of Psychology, 52*, 45-56.
- Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of personality and social psychology, 94*(1), 168.
- SER Diversiteit in bedrijf (2023). *Divers werven en selecteren. Charter document*. Den Haag: SER.
- Shuttleworth-Edwards, A. B. (2016). Generally representative is representative of none: Commentary on the pitfalls of IQ test standardization in multicultural settings. *The Clinical Neuropsychologist, 30*(7), 975-998.

Sitinjak, C. (2023). Literature Study of Cross-Cultural Size Comparisons Against Personality Traits. *Ilomata International Journal of Social Science*, 4(2), 326-335.

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual review of psychology*, 67(1), 415-437.

Steinmayr, R., & Kessels, U. (2017). Good at school= successful on the job? Explaining gender differences in scholastic and vocational success. *Personality and Individual Differences*, 105, 107-115.

Sternberg, R. J., Grigorenko, E., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, 47(1), 1-41.

Stabile, S. J. (2001). The use of personality tests as a hiring tool: is the benefit worth the cost?. *U. Pa. J. Lab. & Emp. L.*, 4, 279.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology*, 69(5), 797.

Suzuki, L. A., & Ahluwalia, M. K. (2003). Gender issues in personality, cognitive, and vocational assessment of women. *Handbook of counseling women*, 119-130.

Schwartz, S. J., Meca, A., Ward, C., Szabó, Á., Benet-Martínez, V., Lorenzo-Blanco, E. I., ... & Zamboanga, B. L. (2019). Biculturalism dynamics: A daily diary study of bicultural identity and psychosocial functioning. *Journal of Applied Developmental Psychology*, 62, 26-37.

Stirratt, M. J., Meyer, I. H., Ouellette, S. C., & Gara, M. A. (2008). Measuring identity multiplicity and intersectionality: Hierarchical classes analysis (HICLAS) of sexual, racial, and gender identities. *Self and Identity*, 7(1), 89-111.

te Nijenhuis, J., de Jong, M. J., Evers, A., & Van Der Flier, H. (2004). Are cognitive differences between immigrant and majority groups diminishing?. *European Journal of Personality*, 18(5), 405-434.

Te Nijenhuis, J., Pesta, B. J., & Fuerst, J. G. R. (2024). General mental ability testing and adverse impact in the United Kingdom: a meta-analysis with more than two million observations. *European Journal of Work and Organizational Psychology*, 33(5), 712-725. <https://doi.org/10.1080/1359432X.2024.2377780>.

Te Nijenhuis, J., Willigers, D., Dragt, J., & van der Flier, H. (2016). The effects of language bias and cultural bias estimated using the method of correlated vectors on a large database of IQ comparisons between native Dutch and ethnic minority immigrants from non-Western countries. *Intelligence*, 54, 117-135.

- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel psychology*, 44(4), 703-742.
- Tett, R. P., & Simonet, D. V. (2021). Applicant faking on personality tests: Good or bad and why should we care?. *Personnel Assessment and Decisions*, 7(1), 2.
- Tett, R. P., Toich, M. J., & Ozkum, S. B. (2021). Trait activation theory: A review of the literature and applications to five lines of personality dynamics research. *Annual Review of Organizational Psychology and Organizational Behavior*, 8(1), 199-233.
- Timmons, K. C. (2020). Pre-Employment personality tests, algorithmic bias, and the Americans with Disabilities Act. *Penn St. L. Rev.*, 125, 389.
- Toth, G., & Siegel, L. S. (2020). A critical evaluation of the IQ-achievement discrepancy based definition of dyslexia. In *Current directions in dyslexia research* (pp. 45-70). Garland Science.
- Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97(3), 499.
- Van Landingham, H., Ellison, R. L., Laique, A., Cladek, A., Khan, H., Gonzalez, C., & Dunn, M. R. (2022). A scoping review of stereotype threat for BIPOC: Cognitive effects and intervention strategies for the field of neuropsychology. *The Clinical Neuropsychologist*, 36(2), 503-522.
- Vazsonyi, A. T., Javakhishvili, M., & Blatny, M. (2022). Does self-control outdo IQ in predicting academic performance?. *Journal of Youth and Adolescence*, 51(3), 499-508.
- Vervliet, T. (2022). *Digitale werving en selectie: algoritmegebruik in het werving- en selectieproces van werkgevers en het bewustzijn van risico's op uitsluiting en discriminatie*. SEO Economisch Onderzoek, Amsterdam.
- Verkuyten, M., & Pouliasi, K. (2002). Biculturalism among older children: Cultural frame switching, attributions, self-identification, and attitudes. *Journal of Cross-Cultural Psychology*, 33(6), 596-609.
- Walton, G. M., Murphy, M. C., & Ryan, A. M. (2015). Stereotype threat in organizations: Implications for equity and performance. *Annual Review of Organizational Psychology and Organizational Behavior*, 2(1), 523-550.
- Wekker, G. & H. Lutz (2001). 'Een hoogvlakte met koude winden. De geschiedenis van het gender- en etniciteitsdenken in Nederland'. M. Botman, N. Jouwe en G. Wekker (red.), *Caleidoscopische visies: de zwarte, migranten- en vluchtelingenvrouwenbeweging in Nederland*. Amsterdam: Koninklijk Instituut voor de Tropen.

Wiggleton-Little, J., & Callender, C. (2023). Screening out neurodiversity. *Kennedy Institute of Ethics Journal*, 33(1), 21-54.

Zell, E., & Lesick, T. L. (2022). Big five personality traits and performance: A quantitative synthesis of 50+ meta-analyses. *Journal of personality*, 90(4), 559-573.

## Colofon

Auteurs: Jolien Geerlings, Hanneke Felten, Emiel Stobbe, Caroline Harnacke, Maxime van de Gevel en Suzan de Winter-Koçak

Met dank aan Marieke Mulder (AWVN), Justin Hoegen Dijkhof (College voor de rechten van de Mens) en Dick Houtzager (NCDR).

Redactie: Joke Martens

Financiering: Ministerie van Sociale Zaken en Werkgelegenheid

Vormgeving: Design Effects, Willem Plaisier

Afbeelding: 123RF

Uitgave: Kennisplatform Inclusief Samenleven

Giessenplein 59C

3522 KE Utrecht

T (030) 230 3260

De publicatie kan gedownload worden via de website van het Kennisplatform Inclusief Samenleven.

© Kennisplatform Inclusief Samenleven, Utrecht, 2024.

Het auteursrecht van deze publicatie berust bij Movisie. Gedeeltelijke overname van teksten is toegestaan, mits daarbij de bron wordt vermeld.

The copyright of this publication rests with Movisie. Partial reproduction of the text is allowed, on condition that the source is mentioned.

## Kennisplatform Inclusief Samenleven

Kennisplatform Inclusief Samenleven doet onderzoek, adviseert en biedt praktische tips en instrumenten over vraagstukken rond integratie, migratie en diversiteit. Daarnaast staat het platform open voor vragen, signalen en meningen en formuleert daar naar beste vermogen een antwoord op.

Deze kennisuitwisseling is bedoeld om een fundamentele bijdrage te leveren aan een pluriforme en stabiele samenleving. Blijf op de hoogte van alle projecten, vragen en antwoorden en andere kennisuitwisseling via [www.kis.nl](http://www.kis.nl), de nieuwsbrief, Twitter en LinkedIn.

Kennisplatform Inclusief Samenleven is een programma van het Verwey-Jonker Instituut en Movisie.

**T** 030 230 32 60

**E** [info@kis.nl](mailto:info@kis.nl)

**I** [www.kis.nl](http://www.kis.nl)

