

REEKS 'WERVING & SELECTIE'

Door de ogen van de beoordelaar: onderzoek naar similar-to-me-effecten op scores van cultureel diverse kandidaten op het assessment center

Lonneke A. L. de Meijer
Erasmus Universiteit Rotterdam
demeijer@essb.eur.nl

Marise Ph. Born
Erasmus Universiteit Rotterdam en North-West University South Africa
m.ph.born@essb.eur.nl

Joran J. Jongerling
Erasmus Universiteit Rotterdam
jongerling@essb.eur.nl

Henk T. van der Molen
Erasmus Universiteit Rotterdam
vandermolen@essb.eur.nl

Samenvatting

McCarthy et al. (2010) en Sacco et al. (2003) vonden geen effect van demografische gelijkheid tussen beoordelaar en kandidaat op beoordelingen van het selectie-interview. Beide studies onderzochten hiërarchisch lineaire modellen (HLM). Gebruikmakend van eenzelfde HLM-procedure is binnen het huidige onderzoek nagegaan wat het effect is van gelijkheid op scores die gegeven zijn aan cultureel diverse kandidaten tijdens de selectie van politieagenten op een ander veelgebruikt instrument, namelijk het assessment center (AC). Meer specifiek onderzochten we het *similar-to-me*-effect van demografische gelijkheid (d.w.z. culturele gelijkheid van de beoordelaar ten aanzien van de kandidaat) en van waargenomen gelijkheid (d.w.z. gelijkheid ten aanzien van specifieke culturele groepen, zoals waargenomen door individuele beoordelaars) op scores op het AC. Demografisch *similar-to-me* bleek gerelateerd te zijn aan *Agency*-scores van autochtoon Nederlandse kandidaten en kandidaten met een migratieachtergrond, maar niet aan

Communion-scores. Waargenomen similar-to-me had geen effect op de beoordelingen van het AC (d.w.z. *Agency* en *Communion*). In zijn algemeenheid bleken de effectgrootten erg klein ($R^2 < .015$). Derhalve kan worden geconcludeerd dat we geen bewijs vonden voor een differentieel effect van gelijkenis op de beoordeling van het AC van cultureel diverse kandidaten, wat bevindingen van McCarthy et al. en Sacco et al. repliceert en uitbreidt.

Abstract

Through the eyes of the assessor: Research into similar-to-me effects on Assessment Center scores of culturally diverse applicants

Research by McCarthy et al. (2010) and Sacco et al. (2003) found no effect of demographic similarity between assessor and applicant on scores given on the employment interview. Both studies used hierarchical linear modeling (HLM). Using a similar HLM approach, the present study explored the effect of similarity on ratings given to native Dutch applicants and applicants with a migration background during police officer selection on another frequently used personnel selection instrument, namely the assessment center (AC). More specifically, we investigated the similar-to-me effect of demographic similarity (i.e., cultural similarity between assessor and applicant), and perceived similarity (i.e., similarity regarding specific cultural groups as considered by the assessor) on the AC ratings. Demographic similar-to-me was found to be related to Agency scores of native applicants and applicants with a migration background, but not to Communion scores. Perceived similar-to-me did not have an effect on ratings given on the AC (i.e., Agency and Communion). In general, the effect sizes were very small ($R^2 < .015$). Therefore, no evidence was found for (dis)similarity differentially affecting evaluations of culturally diverse applicants during the AC, which replicates and extends McCarthy et al.'s and Sacco et al.'s previous research.

Keywords: personnel selection, assessment center, demographic similarity, perceived similarity, cultural diversity

1 Inleiding

Het stimuleren van meer diversiteit op de werkvloer is een belangrijk doel in de geïndustrialiseerde wereld (zie bijv. Cox, 1993; Guillaume et al., 2017). Een strategie in het streven naar meer diversiteit tijdens personeelsselectie is het vermijden van substantiële *adverse impact*. Adverse impact wordt over het

algemeen veroorzaakt door inzet van de intelligentietest (o.a. Murphy, 2002). Binnen de personeelsselectie is op grote schaal gepubliceerd over scores op de intelligentietest en de bestaande verschillen hierin tussen autochtone kandidaten en kandidaten met een migratieachtergrond (zie bijv. Goldstein et al., 2009; Herrnstein & Murray, 1994) en, in enigszins mindere mate, over verschillen in scores op de persoonlijkheidsvragenlijst (zie bijv. De Meijer et al., 2006; Foldes et al., 1998). Mensen met een migratieachtergrond presteren over het algemeen slechter op deze gestandaardiseerde tests. Vanwege de bestaande verschillen in taalvaardigheid die deze verschillen in scores voor een belangrijk deel verklaren (De Meijer et al., 2006; Van den Berg & Van Leest, 1999), is er binnen de personeelsselectie zowel in onderzoek als in de praktijk gezocht naar alternatieven voor geschreven tests (zie bijv. Goldstein et al., 1998) met als doel het stimuleren van meer diversiteit op de werkvloer. Een van die alternatieven is het Assessment Center (AC) (zie bijv. Gaugler et al., 1987). Het huidige onderzoek richt zich op deze alternatieve methode van het AC.

Gerapporteerde verschillen in scores tussen subgroepen op het AC zijn kleiner dan die op bijvoorbeeld de intelligentietest, maar ze zijn nog steeds substantieel. Zo worden in eerder onderzoek scoreverschillen genoemd variërend van 0.02 *SD* tot 0.58 *SD* (zie bijv. Bobko & Roth, 2013; De Meijer et al., 2006), afhankelijk van de aard van de AC-opdracht. Daarbij is een belangrijk kenmerk van het AC dat de beoordeling – anders dan bij de intelligentietest of de persoonlijkheidsvragenlijst – gedaan wordt door een beoordelaar. Omdat bekend is dat interpersoonlijke factoren van invloed zijn op perceptuele processen (Fiske & Taylor, 2016; Klimoski & Donahue, 2001) waaronder beoordelingen, is het interessant om de invloed van gelijkenis tussen beoordelaar en beoordeelde op gegeven beoordelingen nader te onderzoeken. Meer specifiek wordt in het huidige onderzoek bekeken of gelijkenis van de beoordelaar ten aanzien van de kandidaat de manier beïnvloedt waarop beoordelaars de prestaties van autochtoon Nederlandse kandidaten en kandidaten met een migratieachtergrond op het AC beoordelen.

Onder andere McCarthy et al. (2010) en Sacco et al. (2003) onderzochten effecten van gelijkenis. Meer specifiek bestudeerden zij de invloed van demografische gelijkenis tussen beoordelaar en kandidaat op interview-scores onder verschillende culturele groepen in de VS (d.w.z. witte, zwarte, Latino, en Aziatische beoordelaars en kandidaten). Zowel McCarthy et al. (2010) als Sacco et al. (2003) vonden geen effect van demografische gelijkenis op interview-scores. McCarthy et al. (2010) stelden voor dat toekomstig onderzoek zich zou moeten richten op andere instrumenten die gebruikt

worden tijdens personeelsselectie, dat wil zeggen anders dan het interview. Sacco et al. (2003) gaven bovendien aan dat toekomstig onderzoek naar mogelijke effecten van waargenomen gelijkens tussen beoordelaars en kandidaten van belang zou zijn. Het huidige onderzoek komt tegemoet aan zowel het voorstel van McCarthy et al. als het voorstel van Sacco et al. en heeft daarmee twee doelen, namelijk (1) bestudering van de invloed van zowel demografische als waargenomen gelijkens op (2) de beoordelingen met betrekking tot een ander instrument dan het interview, namelijk het AC.

Hoewel er overeenkomsten zijn tussen het selectie-interview en het AC als selectie-instrument – namelijk in beide gevallen wordt een oordeel geveld over gedrag van de kandidaat – zijn er ook duidelijke verschillen aan te wijzen. Anders dan bij het selectie-interview wordt bij het AC namelijk écht gedrag in werk-gerelateerde situaties beoordeeld in plaats van een oordeel over verbale beschrijvingen van gedrag (Dayan et al., 2008). Het AC wordt daarmee gezien als een instrument met een sterke inhouds- en indrukvaliditeit (Thornton & Rupp, 2006). Er is echter ook kritiek op het AC als instrument voor *high stakes*-selectie. De scores van kandidaten op basis van AC-prestaties kunnen namelijk afhankelijk zijn van factoren die niet gecontroleerd kunnen worden, zoals de dynamiek tussen beoordelaars en kandidaten (Fox & Haboucha, 2011). Zo valt te verwachten dat de percepties van de beoordelaar ten aanzien van bijvoorbeeld de attitudes, motivatie of vaardigheden van de beoordeelde van invloed zijn op de uitkomst van perceptuele processen (Fiske & Taylor, 2016; Klimoski & Donahue, 2001). Voor zover bekend, hebben alleen Falk en Fox (2014) effecten van gelijkens op AC-scores onderzocht. Falk en Fox (2014) gebruikten in hun onderzoek een steekproef van Joodse en Arabische beoordelaars en kandidaten in Israël en onderzochten effecten van demografische gelijkens. Zij keken daarbij voornamelijk naar groepsprocessen tijdens het AC en vonden dat de aanwezigheid van een groter aantal Arabische kandidaten in de groep samenhang met hogere AC-scores van de Joodse kandidaten. Zij keken echter niet naar kandidaat-beoordelaar dyades, waar het huidige onderzoek wel op gericht is.

Demografische gelijkens wordt in deze studie opgevat als culturele gelijkens, d.w.z. een (*similar-to-me*-)variabele die aangeeft of de kandidaat en de beoordelaar beiden deel uitmaken van de autochtoon Nederlandse groep of beiden behoren tot een van de groepen met een migratieachtergrond. Eerder onderzoek binnen de personeelsselectie heeft tot nu toe gemengde resultaten laten zien van het effect van demografische gelijkens. Het meeste onderzoek bestudeerde interviewscores. Zowel McCarthy et al. (2010) als Sacco et al. (2003) onderzochten demografische gelijkens en maakten

gebruik van hiërarchisch lineaire modelleren (HLM), waarbij rekening wordt gehouden met de geneste structuur van kandidaten binnen beoordelaars ofwel waarbij rekening wordt gehouden met eventuele onderlinge verschillen tussen beoordelaars. Zij vonden geen bewijs dat culturele gelijkenis een significante rol speelt bij het toekennen van interviewscores aan elk van de groepen kandidaten. Buckley et al. (2007) en McFarland et al. (2004) gebruikten geen HLM maar variantieanalyses om zwarte en witte interviewers en kandidaten in de V.S. te onderzoeken. Zij lieten zien dat zwarte beoordelaars positievere beoordelingen gaven aan zwarte kandidaten dan aan witte kandidaten, maar alleen als de compositie van het panel van beoordelaars voornamelijk uit zwarte beoordelaars bestond.

Waargenomen gelijkenis wordt in het huidige onderzoek opgevat als waargenomen gelijkenis tussen groepen, d.w.z. een (*similar-to-me*-)variabele die de waargenomen gelijkenis van de beoordelaar aangeeft ten aanzien van een gemiddeld of prototypisch lid van een bepaalde culturele groep. Wanneer waargenomen gelijkenis onderzocht wordt, gaat het niet zozeer om objectieve demografische kenmerken, maar eerder om de percepties van gelijkenis, waaronder gelijkenis in minder tastbare attributies, zoals waarden, overtuigingen en persoonlijkheid. Het meeste onderzoek naar waargenomen gelijkenis had betrekking op werknemers en hun prestaties (Ensher & Murphy, 1997; Lankau et al., 2005; Strauss et al., 2001; Turban et al., 2002). Voor zover bekend zijn er geen studies geweest die, net als de studies van McCarthy et al. (2010) en Sacco et al. (2003), gebruik hebben gemaakt van HLM als analysemethode bij onderzoek naar het effect van waargenomen gelijkenis. Desalniettemin hebben studies die terecht gebruik hebben gemaakt van ANOVA of regressieanalyse als techniek om data te bestuderen zonder geneste structuur (bijv. Ensher & Murphy, 1997; Lankau et al., 2005), over het algemeen steun gevonden voor het idee dat waargenomen gelijkenis samenhangt met relevante uitkomstmaten (bijvoorbeeld de kwaliteit van mentoring op het werk).

Bovenstaande studies naar demografische en waargenomen gelijkenis toetsten de gelijkenis-hypothese, waarbij de veronderstelling was dat gelijkenis leidt tot positievere beoordelingen. Deze verwachting is afgeleid van het gelijkenis-aantrekking-paradigma (*Similarity-Attraction-Paradigm*; Byrne, 1971). Het gelijkenis-aantrekking-paradigma heeft zijn oorsprong in de sociale identiteitstheorie (Tajfel, 1982; Turner, 1987) en de zelf-categorisatietheorie (Turner, 1987), die inhouden dat aspecten van het zelfbeeld van iemand afgeleid worden van de sociale categorieën waartoe deze persoon zich vindt behoren (bijv. de culturele groep, geslacht). Sociale identiteit wordt gezien als belangrijk voor iemands zelfvertrouwen. Bijvoorbeeld, wanneer

een persoon tot een sociale categorie behoort waarvan de leden succesvol zijn, stimuleert dit succes het zelfbeeld van deze persoon.

Sociale identiteiten en zelf-categorisaties van mensen worden opgebouwd aan de hand van demografische kenmerken. Aangenomen wordt dat demografische gelijkens tot positievere attitudes en gedragingen ten aanzien van de ander leidt, omdat dit het positieve zelfbeeld stimuleert. Demografische verschillen daarentegen zullen eerder een negatief effect hebben op de attitudes en gedragingen ten aanzien van de ander. In lijn met dit idee verwachten we in de huidige studie het volgende:

Hypothese 1: Demografische gelijkens van de beoordelaar ten aanzien van de kandidaat (demografisch *similar-to-me*) zal leiden tot hogere beoordelingen op het AC.

Verschillende onderzoekers hebben een sterker effect gevonden voor waargenomen dan voor demografische gelijkens, bijvoorbeeld op de kwaliteit van mentoringrelaties (zie bijv. Ensher et al., 2002) en op de beoordeling van personeel (zie bijv. Strauss et al., 2001). Ferris en Judge (1991) suggereren dat een reden voor het vinden van sterkere effecten van percepties van gelijkens zou kunnen zijn dat mensen reageren op basis van *percepties* van de werkelijkheid, en niet op basis van de werkelijkheid zelf. Deze redenering volgend verwachten wij het volgende:

Hypothese 2: Hoe meer een beoordelaar een gemiddelde, prototypische kandidaat van een bepaalde culturele groep als gelijk aan zichzelf waarneemt (waargenomen *similar-to-me*), hoe hoger de kandidaat die tot deze groep behoort, beoordeeld wordt op het AC.

Een grote dataset van de selectieprocedure van politieagenten in Nederland ($N = 27746$) kon worden gebruikt, waaronder data van kandidaten met een autochtoon Nederlandse achtergrond en kandidaten afkomstig uit de vier grootste groepen met een migratieachtergrond in Nederland, namelijk kandidaten met een Surinaamse achtergrond, een voormalig-Antilliaanse/Arubaanse, een Turkse of een Marokkaanse achtergrond (iemand met een migratieachtergrond is in navolging van de CBS-definitie (2016) iemand van wie ten minste één ouder in het buitenland is geboren). Gezamenlijk vormen deze groepen met een migratieachtergrond ongeveer 7.6% van de Nederlandse bevolking (CBS, 2019). In totaal heeft 23.6% van de Nederlandse bevolking een migratieachtergrond en 13.3% een niet-Westerse migratieachtergrond (CBS, 2019). Deze cijfers zijn al enkele jaren min of meer

constant. Ter vergelijking, in 2005 had 7.0% van de Nederlandse bevolking een Surinaamse, voormalig-Antilliaanse/Arubaanse, Turkse of Marokkaanse achtergrond (CBS, 2019).

2 Methode

2.1 Deelnemers

De gegevens zijn afkomstig van 27746 kandidaten die tussen september 2001 en februari 2007 solliciteerden naar een positie bij de Nederlandse Politieacademie. Daarvan hadden 3089 kandidaten (11%) een Surinaamse achtergrond, een voormalig-Antilliaanse/Arubaanse, een Turkse of een Marokkaanse achtergrond. Data van 147 beoordelaars (84% vrouw, $n = 12$ beoordelaars hadden een migratieachtergrond) werden gebruikt om het effect van *demografische*, d.w.z. *culturele*, *gelijkenis* te bestuderen op scores verkregen op het AC (Hypothese 1). De relatie tussen *waargenomen gelijkenis tussen groepen* en scores op het AC (Hypothese 2) werd onderzocht met een deel-steekproef van 15 beoordelaars (80% vrouw, 93% autochtoon Nederlands) die een waargenomen-gelijkenisvragenlijst hadden ingevuld. Deze specifieke deel-steekproef van beoordelaars werd gevraagd om de vragenlijst in te vullen, omdat zij allen een voldoende grote groep autochtoon Nederlandse kandidaten hadden beoordeeld én een voldoende grote groep kandidaten met een migratieachtergrond hadden beoordeeld (zie Tabel 1 voor een gedetailleerd overzicht van de kandidaat-beoordelaar-dyades).

2.2 Procedure

De gegevens van de beoordelaars en de gegevens van de kandidaten en hun beoordelingen op het AC zijn tussen 2001 en 2007 verzameld door de Nederlandse Politieacademie. Een geanonimiseerde dataset van deze periode is aan de auteurs ter beschikking gesteld voor wetenschappelijk onderzoek. Door de eerste auteur is daarnaast een waargenomen-gelijkenisvragenlijst uitgezet bij een deel van de beoordelaars. Deze beoordelaars hadden elk een voldoende grote groep kandidaten beoordeeld. Het invullen van de vragenlijst was op vrijwillige basis. Na uitleg over het doel van het onderzoek hebben alle beoordelaars die benaderd werden, de vragenlijst ingevuld.

Tabel 1 Gemiddelden en standaarddeviaties van beoordelingen en steekproefgrootten van elke kandidaat-beoordelaar-combinatie

Migratieachtergrond kandidaat	Demografische gelijkens		Waargenomen gelikens
	Beoordelaars		Beoordelaars
	Autochtoon	Migratie- achtergrond	Totaal
Autochtoon			
$M (SD)_{Agency}$	4.08 (0.94)	4.01 (0.95)	4.17 (0.96)
$M (SD)_{Communion}$	3.96 (1.16)	3.97 (1.17)	4.12 (1.15)
Kandidaat n	20995	2901	5390
Beoordelaar n	135	12	15
Voormalig-Antilliaans			
$M (SD)_{Agency}$	3.77 (1.00)	3.90 (0.98)	3.88 (1.04)
$M (SD)_{Communion}$	3.74 (1.19)	4.13 (1.14)	4.05 (1.04)
Kandidaat n	172	30	47
Beoordelaar n	58	7	9
Marokkaans			
$M (SD)_{Agency}$	3.69 (0.90)	3.83 (0.95)	3.86 (0.81)
$M (SD)_{Communion}$	3.57 (1.19)	3.67 (1.33)	3.76 (1.19)
Kandidaat n	413	62	114
Beoordelaar n	85	6	13
Surinaams			
$M (SD)_{Agency}$	3.79 (0.97)	3.94 (0.98)	4.02 (0.88)
$M (SD)_{Communion}$	3.78 (1.22)	3.85 (1.22)	3.98 (1.20)
Kandidaat n	521	81	136
Beoordelaar n	96	8	14
Turks			
$M (SD)_{Agency}$	3.56 (0.99)	3.47 (0.96)	3.75 (0.92)
$M (SD)_{Communion}$	3.54 (1.22)	3.57 (1.12)	3.64 (1.17)
Kandidaat n	841	108	225
Beoordelaar n	105	8	15

2.3 Meetinstrumenten

Assessment Center- (AC)-opdracht

De AC-opdracht is ontwikkeld met als doel authentiek gedrag van kandidaten te ontlokken. Omdat de kandidaten geen voorkennis hebben of zouden kunnen hebben van relevant werk als politieagent, is de AC-opdracht – dat wil zeggen, een rollenspel – geen directe simulatie van de taken van een politieagent. Twee verschillende rollenspelen werden gebruikt; beide betroffen een bepaalde probleemsituatie in een min of meer publieke ruimte. Bij het

ene rollenspel speelde een woordenwisseling op de markt. Bij het andere ging het om een probleem op een vliegveld. De kandidaten werden beoordeeld op hun vaardigheid te de-escaleren en hulp te bieden aan diegenen die dat nodig hadden. De rollenspelen zijn onderling uitwisselbaar en kandidaten namen slechts deel aan één van beide rollenspelen. Er waren steeds drie personen betrokken bij elk rollenspel, namelijk een acteur, een beoordelaar en de kandidaat. De beoordelaar en de acteur beoordeelden en speelden om beurten in opeenvolgende rollenspelen.

De rollenspelen verliepen alle volgens eenzelfde gestructureerde opzet, die in het onderstaande uiteengezet zal worden. De voorbereiding op het rollenspel start op het moment dat de beoordelaar samen met de kandidaat naar een kamer loopt, waar de kandidaat aan de hand van een geschreven instructie 15 minuten de tijd heeft om zich voor te bereiden op het rollenspel. De beoordelaar verschaft geen aanvullende informatie. De kandidaat bestudeert tijdens de gegeven 15 minuten de uitleg over de fictieve situatie, inclusief informatie over de rol van de kandidaat en de fictieve situatie die zich voorafgaand aan het rollenspel had afgespeeld. Na 15 minuten wordt de kandidaat opgehaald en tot aan de deur van de kamer geleid waar het rollenspel plaatsvindt. De enige twee dingen die op dat moment nog door de beoordelaar worden gezegd, zijn: "Heb je op dit moment nog vragen?" en "De simulatie start zodra je de kamer binnengaat". De beoordelaar gaat vervolgens de kamer binnen; de kandidaat dient dan nog even buiten te blijven staan. In de kamer is tevens een acteur aanwezig. De beoordeling van het rollenspel start op het moment dat de kandidaat de deur opent en het rollenspel van start gaat. De acteur volgt een gedetailleerd script over hoe te interacteren met de kandidaat. De beoordelaar, die op dat moment niet deelneemt aan het rollenspel, observeert, luistert naar de kandidaat en maakt notities van wat de kandidaat zegt en doet. Het rollenspel duurt 15 minuten.

Aan het einde van de AC-opdracht wordt de kandidaat door de beoordelaar op de volgende zeven dimensies beoordeeld: Communicatievaardigheden, Sociale vaardigheden, Empathie, Initiatief, Stressbestendigheid, Autoriteit en Besluitvaardigheid. De beoordeling vindt plaats aan de hand van een 7-punt Likertschaal variërend van 1 (extreem zwak) tot 7 (excellent). Elke dimensie wordt beoordeeld aan de hand van een gestructureerd formulier met zo'n vier items per dimensie en daarbij uitgeschreven schaal-ankers per Likertpunt. De beoordelaar stelt zijn beoordeling alleen op. In uitzonderlijke gevallen van twijfel gaat de beoordelaar in gesprek met de acteur om tot een beoordeling te komen. Inter-beoordelaarsbetrouwbaarheden varieerden van .82 tot .88 (vastgesteld op basis van een deel-steekproef van $N_{\text{beoordelaar}}$

= 198 en $N_{\text{acteur}} = 198$). Principale componentenanalyse met varimaxrotatie toonde twee dimensies, *Agency* en *Communion* genaamd (in navolging van Wiggins & Trapnell, 1996), die samen 77% van de variantie verklaarden. De dimensies Autoriteit, Besluitvaardigheid, Initiatief, Communicatievaardigheden, en Stressbestendigheid laadden op *Agency* en als maat hiervoor is het gemiddelde gebruikt ($\bar{r} = .59, \alpha = .87$). De dimensies Sociale vaardigheden en Empathie laadden op *Communion*. Het gemiddelde is gebruikt als maat voor *Communion* ($\bar{r} = .77, \alpha = .87$). In termen van gedrag correspondeert *Agency* met de taak 'handhaving van de rechtsorde' uit artikel 2 van de Nederlandse politiewet. *Communion* correspondeert met de 'hulpverleningstaak' uit artikel 2. De betrouwbaarheid van het verschil (r_{diff}) tussen de scores op *Agency* en *Communion* is .78.

Vragenlijst Waargenomen Gelijkenis

Beoordelaars die een voldoende grote groep van autochtoon Nederlandse kandidaten én kandidaten met een migratieachtergrond hadden beoordeeld, werden gevraagd om op vrijwillige basis een *paper-and-pencil*-vragenlijst met betrekking tot waargenomen gelijkenis in te vullen. Zij werden gevraagd de uitgereikte vragenlijst na invullen te overhandigen aan de eerste onderzoeker. De Vragenlijst Waargenomen Gelijkenis is gebaseerd op een schaal van McCroskey et al. uit 1975. Bij deze schaal, die bestaat uit 17 items, geven beoordelaars aan in welke mate zij een bepaalde culturele groep (in dit geval groepen met een voormalig-Nederlands Antilliaanse, een Surinaamse, een Turkse of een Marokkaanse afkomst) gelijk zien aan henzelf op vier aspecten, namelijk attituden, waarden, fysieke verschijning en achtergrond, met een gemiddeld lid van de groep in gedachten. Elk van de 17 items moest dus worden ingevuld voor elk van de vier culturele groepen. Een voorbeelditem is 'Ik ben van mening dat Turkse mensen dezelfde normen en waarden hebben als ik heb' (Likertschaal van 1 tot 7). Hetzelfde item moest ook worden ingevuld voor de voormalig-Nederlands Antilliaanse groep, voor de Surinaamse en voor de Marokkaanse groep. Voor elke culturele groep zijn uiteindelijk de scores op de vier aspecten gemiddeld, omdat de intercorrelaties tussen de percepties van gelijkenis vrij hoog waren ($\bar{r}_{\text{Antilliaans}} = .70, \bar{r}_{\text{Marokkaans}} = .64, \bar{r}_{\text{Surinaams}} = .67, \bar{r}_{\text{Turks}} = .67; \alpha_{\text{Antilliaans}} = .91, \alpha_{\text{Marokkaans}} = .92, \alpha_{\text{Surinaams}} = .88, \alpha_{\text{Turks}} = .90$).

Waargenomen Integratie

Aanvullend op de Vragenlijst Waargenomen Gelijkenis werd gevraagd naar de waargenomen integratie in de samenleving. Onze groep van kandidaten met een migratieachtergrond bestond uit subgroepen met verschillende

culturele achtergronden. We wilden onderzoeken wat het effect was van demografische en waargenomen gelijkens voor deze verschillende subgroepen, gedifferentieerd naar de waargenomen mate van integratie in de maatschappij van een bepaalde groep. We hielden hierbij de definitie van integratie aan, zoals Berry (1997, p. 9) deze gebruikte: 'Als een persoon geïntegreerd is in de maatschappij, is er zowel een interesse in het behouden van de eigen oorspronkelijke cultuur als een gerichtheid op dagelijkse interactie met andere groepen.' Met andere woorden, een bepaalde mate van behoud van de oorspronkelijke culturele identiteit blijft in dat geval aanwezig terwijl men ook zoekt naar deelname aan een groter sociaal netwerk in de dominante cultuur (Segall et al., 1999). Waargenomen integratie in de samenleving werd gemeten door beoordelaars te vragen in welke mate zij een gemiddeld lid van een bepaalde culturele groep (in dit geval groepen met een voormalig-Nederlands Antilliaanse, een Surinaamse, Turkse of Marokkaanse afkomst) als geïntegreerd in de Nederlandse samenleving zagen (op een 4-punt Likertschaal).

Eerder onderzoek bestudeerde reeds de verschillen tussen groepen met een migratieachtergrond. In Groot-Brittannië bijvoorbeeld toonden Carmichael en Woods (2000) en Modood (1997) aan dat er verschillen bestaan in de kans op een baan tussen verschillende culturele groepen, waarbij de ene groep met een migratieachtergrond meer benadeeld wordt dan de andere groep met een migratieachtergrond. Vanuit een ietwat andere invalshoek zijn in Nederland verschillende studies gedaan naar de integratie-hiërarchie van verschillende culturele groepen (zie bijv. Stupar et al., 2014; Verkuyten et al., 1996; Weijters & Scheepers, 2003). In deze hiërarchie van culturele groepen staan niet-Nederlandse Europese groepen bovenaan, gevolgd door voormalig Nederlands-koloniale groepen en als laatste groepen met een Arabische achtergrond. De grootste groepen met een migratieachtergrond in Nederland zijn afkomstig uit de voormalige Nederlandse Antillen en Aruba, en uit Marokko, Suriname en Turkije. De voormalige Nederlandse Antillen met Aruba en Suriname zijn vroegere Nederlandse koloniën. Personen met een Marokkaanse en Turkse achtergrond hebben veelal een Arabisch culturele achtergrond.

Een verklaring waarom de groepen met een Turkse of Marokkaanse achtergrond worden gezien als minder geïntegreerd in de Nederlandse samenleving dan groepen met een Nederlands Antilliaanse, Arubaanse of Surinaamse achtergrond, kan gevonden worden in de relatief grote culturele verschillen (Stupar et al., 2014). Zo veronderstelden onder anderen Stupar et al. (2014) dat de Arabische cultuur meer verbonden is met het islamitisch geloof en vanuit dit geloof andere waarden aanhangt met betrekking tot

familie, het huwelijk en de rol van de vrouw, dan de waarden van Surinaamse en voormalig-Antilliaanse en Arubaanse groepen. Door de verbondenheid met het christendom van deze laatstgenoemde groepen worden zij veelal als meer geïntegreerd ervaren door de autochtoon Nederlandse groep. Een andere verklaring voor de gevonden integratie-hiërarchie is het integratie-beleid van arbeidsmigranten dat in de jaren zestig van de vorige eeuw in Nederland uitgevoerd werd (Jennissen, 2009). Aan de ene kant stimuleerde de Nederlandse regering Turkse en Marokkaanse immigranten om hun oorspronkelijke cultuur te behouden, omdat men verwachtte dat ze op een zeker moment zouden repatriëren. Aan de andere kant waren Surinaamse en Antilliaanse groepen, vanwege het koloniale verleden, al meer bekend met de Nederlandse taal en cultuur op het moment van aankomst in Nederland (Stupar et al., 2014).

2.4 Initiële analyses

Initiële analyses werden uitgevoerd om te onderzoeken in hoeverre de vier culturele groepen in onze steekproef waargenomen worden als geïntegreerd in de Nederlandse samenleving. Een Chi-kwadraattest liet zien dat de vier culturele groepen als niet gelijk-geïntegreerd werden waargenomen ($\chi^2 (df=3) = 24.26, p < .001$). Uit de resultaten (zie Bijlage) volgde dat de Surinaamse en de voormalig Antilliaanse groepen als meest geïntegreerd werden gezien, gevolgd door de Turkse en Marokkaanse groepen. In het verdere onderzoek naar demografische en waargenomen gelijkheid is daarom – vergelijkbaar met eerder onderzoek in de Nederlandse context (zie Stupar et al., 2014) – gedifferentieerd naar drie groepen kandidaten, namelijk de autochtoon Nederlandse groep, de Surinaams/Antilliaanse groep en de Turkse/Marokkaanse groep.

2.5 Analyses

Om het effect van demografische gelijkheid (Hypothese 1) en waargenomen gelijkheid (Hypothese 2) te onderzoeken op scores die gegeven worden op het AC, pasten we verschillende multilevel *structural equation*-modellen toe op de data door middel van het programma Mplus (Muthén & Muthén, 1998-2011). Deze modellen laten onderzoek toe naar complexe relaties tussen verschillende variabelen tegelijk (bijvoorbeeld met meerdere afhankelijke variabelen), terwijl rekening wordt gehouden met de afhankelijkheid van gegevens zoals geneste datastructuren (in dit onderzoek: kandidaten genest binnen beoordelaars).

Op het eerste ofwel kandidaat-niveau zijn demografische en waargenomen gelijkheid gemodelleerd als voorspellers van scores op de afhankelijke

variabelen, namelijk de scores op *Agency* en *Communion* van het AC, terwijl gecontroleerd werd voor migratieachtergrond van de kandidaat. Migratieachtergrond van de kandidaat werd gecodeerd aan de hand van de drie eerdergenoemde groepen, namelijk de autochtoon Nederlandse groep, de Surinaams/voormalig-Antilliaanse groep en de Turks/Marokkaanse groep. Deze drie groepen werden aan het model toegevoegd als dummy-variabelen, met een dummy voor de Surinaams/voormalig-Antilliaanse groep en een dummy voor de Turks/Marokkaanse groep.

Demografische gelijkens

Demografische gelijkens werd geoperationaliseerd als een variabele die aangeeft of de kandidaat en de beoordelaar beiden autochtoon Nederlands zijn of beiden een migratieachtergrond hebben. Als de kandidaat en de beoordelaar beiden autochtoon Nederlands waren of beiden een migratieachtergrond hadden, werd demografische gelijkens gescoord als 0. Een score van 1 gaf aan dat de kandidaat autochtoon Nederlands was én de beoordelaar een migratieachtergrond had of andersom. Op het tweede of beoordelaar-niveau werd de migratieachtergrond van de beoordelaar opgenomen als voorspeller van *Agency* en *Communion*. De migratieachtergrond van de beoordelaar werd gescoord als 0 als de beoordelaar autochtoon Nederlands was en als 1 als de beoordelaar een migratieachtergrond had. Vanwege het kleine aantal beoordelaars met een migratieachtergrond ($n = 12$) was het niet mogelijk om verder te differentiëren binnen deze groep.

Waargenomen gelijkens

Elke kandidaat kreeg een waargenomen-gelijkenisscore die zijn of haar beoordelaar gaf aan een gemiddeld lid van de specifieke culturele groep waartoe de kandidaat behoorde. Alle continue variabelen, d.w.z. waargenomen gelijkens, *Agency* en *Communion*, werden gestandaardiseerd (zie Raudenbush & Bryk, 2002).

3 Resultaten

3.1 Initiële bevindingen

Om een beeld te krijgen van de verschillen in scores tussen de autochtoon Nederlandse groep kandidaten en de groep met een migratieachtergrond zijn de parameterschattingen van de dummy-variabelen relevant. Deze waren gemaakt voor twee groepen kandidaten namelijk de Surinaams/voormalig-Antilliaanse groep en de Turks/Marokkaanse groep (zie Tabellen 3 en 5). Deze

parameter-schattingen geven weer of er een significant verschil in scores ten aanzien van de autochtoon Nederlandse groep bestaat. We verwijzen graag naar een studie De Meijer et al. (2006) voor een gedetailleerder beeld van de bestaande scoreverschillen tussen verschillende culturele groepen in Nederland op het AC en, meer specifiek, van de effectgrootten. In die studie zijn eerste-generatie- en tweede-generatie-minderheidsgroepen vergeleken met de autochtoon Nederlandse groep in een andere maar vergelijkbare steekproef van 13526 kandidaten. De resultaten lieten zien dat scoreverschillen op het AC tussen de autochtone groep en groepen met een migratieachtergrond grofweg gelijk waren aan bevindingen uit Noord-Amerikaans onderzoek en varieerden tussen 0.02 *SD* en 0.59 *SD*, in de verwachte richting. Verder lieten de resultaten zien dat de scoreverschillen tussen de autochtoon Nederlandse groep en de Surinaamse en de voormalig-Antilliaanse groep iets kleiner waren dan tussen de autochtoon Nederlandse groep en de Turkse en de Marokkaanse groep.

3.2 Demografische gelijkens

In de *demografische gelijkens*-hypothese stelden we dat culturele gelijkens van de beoordelaar ten aanzien van de kandidaat zou leiden tot hogere scores (Hypothese 1). Om dit te toetsen hebben we de relaties onderzocht tussen de scores op het AC – namelijk op *Agency* en *Communion* – aan de ene kant, en de predictoren migratieachtergrond van de kandidaat en van de beoordelaar en demografische gelijkens van de beoordelaar ten aanzien van de kandidaat aan de andere kant. Om te bepalen of multilevel-analyse nodig was, werden eerst twee benchmark-modellen getoetst (zie Tabel 2). Het eerste model was het *nul*-model dat geen enkele variabele op Niveau-2 modelleert. Indien dit *nul*-model past, kunnen we aannemen dat er geen Niveau-2-structuur in de data aanwezig is. Dit zou betekenen dat multilevel-analyse niet nodig is (Hox, 2010). In ons *nul*-model voorspelden we de afhankelijke variabelen *Agency* en *Communion* aan de hand van de onafhankelijke variabelen op kandidaat-niveau (Niveau-1): migratieachtergrond van de kandidaat (onderscheiden naar de autochtoon Nederlandse groep, de Surinaams/voormalig-Antilliaanse groep en de Turkse/Marokkaanse groep) en demografische gelijkens, terwijl alle variabelen op Niveau 2 (d.w.z. de varianties in *Agency* en *Communion*) op 0 gefixeerd werden. Het tweede model dat getoetst werd, was het *independence*-model. Het *independence*-model had hetzelfde Niveau-1-model als het *nul*-model, maar – anders dan het *nul*-model – werd hierbij variantie op Niveau-2 toegelaten van de afhankelijke variabelen *Agency* en *Communion* (maar geen covariantie). Als dit model past bij de data, kan worden aangenomen dat er variantie op Niveau-2 bestaat en, als

Tabel 2 Demografische gelijkenis: twee benchmarkmodellen

Model	χ^2	df
<i>Nul</i> -model	895.26*	3
<i>Independence</i> -model	924.64*	1

NB. Het *nul*-model modelleert alleen variabelen op Niveau-1 (d.w.z. *Agency* en *Communion* zoals gemeten met het AC, migratieachtergrond van de kandidaat, en demografische gelijkenis tussen beoordelaar en kandidaat). Het *independence*-model laat ook variantie op Niveau-2 toe, maar geen covariantie.

* $p < .001$

gevolg hiervan, dat hiërarchische modellering nodig is, maar dat er geen structurele relaties aanwezig zijn op Niveau-2 (Hox, 2010). De resultaten in Tabel 2 laten zien dat zowel het *nul*- als het *independence*-model niet goed bij de data passen. Dit houdt in dat niet alleen hiërarchische modellering nodig is, maar dat het tevens nodig is om structurele relaties op Niveau-2 te specificeren (Hox, 2010), d.w.z. dat het nodig is om in het model te controleren voor culturele achtergrond van de beoordelaar.

In het derde model werd, naast migratieachtergrond van de kandidaat en demografische gelijkenis van de beoordelaar ten aanzien van de kandidaat op Niveau-1, de Niveau-2-predictor migratieachtergrond van de beoordelaar opgenomen. De resultaten van dit derde model in Tabel 3 laten in eerste instantie een negatieve relatie zien tussen migratieachtergrond van de kandidaat en de twee uitkomstmaten. Dit betekent dat kandidaten met een migratieachtergrond een significant lagere score ontvingen op de dimensies gemeten met het AC (d.w.z. *Agency* en *Communion*). In de tweede plaats bleek dat de migratieachtergrond van de beoordelaar geen invloed had op AC-scores. Wel bleek er een significant negatieve relatie te bestaan tussen demografische gelijkenis van de beoordelaar ten aanzien van de kandidaat en de score op *Agency*, wat betekent dat culturele gelijkenis van de beoordelaar ten aanzien van de kandidaat resulteert in een hogere score op *Agency*. Geen relatie werd gevonden tussen demografische gelijkenis en scores op *Communion*. Wanneer echter een steekproef groot is, is de kans op het vinden van een significant effect hoog, ook als de daadwerkelijke effectgrootten klein zijn. Om deze reden zijn tevens effectgrootten berekend en opgenomen bij de resultaten. Hieruit blijkt dat wanneer het percentage verklaarde variantie in overweging wordt genomen – dit was 1.4% van de Niveau-1-variantie in *Agency* (en 0.6% in *Communion*) – slechts een klein deel van de variantie in scoreverschillen op het AC (Cohen, 1988) verklaard wordt door demografische gelijkenis van de beoordelaar ten aanzien van de kandidaat. De conclusie is dan ook dat we slechts minimale steun vonden voor Hypothese 1.

Tabel 3 Demografische gelijkenis: resultaten van het model met Niveau-1 predictoren migratieachtergrond van de kandidaat en demografische gelijkenis, en de Niveau-2 predictor migratieachtergrond van de beoordelaar

Vergelijking groepen kandidaten	Parameter-schattingen B (SE)	R^2
Niveau-1		
<i>Agency</i>		0.014
1 Surinaams/voormalig-Antilliaans	-0.18** (0.04)	}
2 Turks/Marokkaans	-0.44** (0.03)	
3 Demografische gelijkenis	-0.06* (0.03)	0.001
<i>Communion</i>		0.006
1 Surinaams/voormalig-Antilliaans	-0.08* (0.04)	}
2 Turks/Marokkaans	-0.31** (0.03)	
3 Demografische gelijkenis	-0.03, <i>ns</i> (0.02)	0.000
Niveau-2		
<i>Agency</i>		0.000
1 Migratieachtergrond beoordelaar	0.00, <i>ns</i> (0.06)	
<i>Communion</i>		0.000
1 Migratieachtergrond beoordelaar	0.06, <i>ns</i> (0.05)	

NB. Een negatieve Niveau-1 parameter-schatting in rij 1 en 2 betekent dat kandidaten met een migratieachtergrond lagere AC-scores ontvangen (d.w.z. op *Agency* en *Communion*) dan autochtone kandidaten. Een negatieve parameter-schatting in rij 3 betekent dat culturele gelijkenis tussen kandidaat en beoordelaar resulteert in een hogere score (na correctie voor migratieachtergrond van de beoordelaar).

* $p < .05$, ** $p < .001$, *ns* betekent niet-significant

3.3 Waargenomen gelijkenis

Met betrekking tot *waargenomen gelijkenis* werd aangenomen dat hoe meer een beoordelaar een gemiddeld persoon van een bepaalde culturele groep als gelijk aan zichzelf ziet, hoe hoger een kandidaat beoordeeld wordt die behoort tot deze culturele groep (Hypothese 2). Deze hypothese werd op een overeenkomstige manier getoetst als de demografische-gelijkenis-hypothese. We onderzochten de relaties tussen scores op het AC – d.w.z. op *Agency* en *Communion* – aan de ene kant en de predictoren migratieachtergrond van de kandidaat (onderscheiden naar de autochtoon Nederlandse groep, de Surinaams/voormalig-Antilliaanse groep en de Turkse/Marokkaanse groep), migratieachtergrond van de beoordelaar en waargenomen gelijkenis aan de andere kant. Om na te gaan of multilevel-analyse nodig was, werden wederom twee benchmark-modellen getoetst (zie Tabel 4).

De resultaten in Tabel 4 laten zien dat zowel het *nul-* als het *independence-*model niet bij de data paste. Dit betekent dat multilevel-analyse nodig is én dat structurele relaties op Niveau-2 gespecificeerd moeten worden

Tabel 4 Waargenomen gelijkenis: twee benchmarkmodellen

Model	χ^2	df
Nul-model	241.02*	3
Independence-model	22.64*	1

NB. Het *nul*-model modelleert alleen variabelen op Niveau-1 (d.w.z. *Agency* en *Communion* zoals gemeten met het AC, migratieachtergrond van de kandidaat, en waargenomen gelijkenis van kandidaten). Het *independence*-model laat ook variantie op Niveau-2 toe, maar geen covariantie.

* $p < .001$

Tabel 5 Waargenomen gelijkenis: resultaten van het model met Niveau-1 predictoren migratieachtergrond van de kandidaat en waargenomen gelijkenis, en de Niveau-2 covariantie tussen *Agency* en *Communion*

Vergelijking groepen kandidaten	Parameter-schattingen B (SE)	R^2
Niveau-1		
<i>Agency</i>		0.008
1 Surinaams/voormalig-Antilliaans	-0.09, <i>ns</i> (0.05)	0.007
2 Turks/Marokkaans	-0.37* (0.04)	
3 Waargenomen gelijkenis	-0.02, <i>ns</i> (0.02)	
<i>Communion</i>		0.007
1 Surinaams/voormalig-Antilliaans	-0.03, <i>ns</i> (0.06)	0.007
2 Turks/Marokkaans	-0.36* (0.07)	
3 Waargenomen gelijkenis	-0.01, <i>ns</i> (0.04)	

NB. Een negatieve parameter-schatting in rij 2 betekent dat kandidaten met een migratieachtergrond lagere AC-scores ontvangen (d.w.z. op *Agency* en *Communion*) dan kandidaten met een autochtoon Nederlandse achtergrond.

* $p < .001$, *ns* betekent niet-significant

(Hox, 2010). Als gevolg hiervan werd een derde model gespecificeerd met daarin de Niveau-2-predictor migratieachtergrond van de beoordelaar, naast migratieachtergrond van de kandidaat en waargenomen gelijkenis als Niveau-1-predictoren zoals in het *nul*- en *independence*-model. Dit model paste echter niet goed, waarschijnlijk vanwege het relatief grote aantal te schatten parameters ten aanzien van de steekproefgrootte op Niveau-2 (d.w.z. 16 te schatten parameters in vergelijking met 15 beoordelaars (Niveau-2-steekproefgrootte)). Om die reden werd besloten om migratieachtergrond van de beoordelaar achterwege te laten in de analyse en *Agency* en *Communion* alleen te laten covariëren op Niveau-2. De resultaten in Tabel 5 met betrekking tot dit laatste model laten zien dat de Turks/Marokkaanse groep significant lager scoorde op de AC-dimensies *Agency* en *Communion* dan de autochtoon Nederlandse groep, terwijl de Surinaams/voormalig-Antilliaanse

groep niet significant verschilde van de autochtone groep. Daarbij bleek waargenomen gelijkens niet van invloed te zijn op de scores die gegeven werden op het AC. Als geheel voorspelde dit model 0.8% en 0.7% van de Niveau-1-variantie van respectievelijk *Agency* en *Communion*. Daarmee werd er geen steun gevonden voor Hypothese 2.

Samengevat kan worden geconcludeerd dat demografische gelijkens van de beoordelaar ten aanzien van de kandidaat een significante relatie heeft met de AC-dimensie *Agency* (maar er werd geen significante relatie gevonden met de AC-dimensie *Communion*), terwijl waargenomen gelijkens ten aanzien van de kandidaat niet gerelateerd was aan de gegeven scores op het AC.

4 Discussie

Zowel vanuit het oogpunt van inclusiviteit als vanuit het idee dat teams met meer diversiteit effectiever zijn, wordt diversiteit op de werkvloer meer en meer gestimuleerd. Een strategie in het streven naar meer diversiteit is het vermijden van *adverse impact* tijdens personeelsselectie. Adverse impact wordt over het algemeen veroorzaakt door het inzetten van de intelligentietest (o.a. Murphy, 2002). Deze strategie bestaat uit het gebruik van een serie indrukvalide selectie-instrumenten, anders dan de intelligentietest, zonder dat de voorspellende waarde van de selectieprocedure hieronder lijdt. Een voorbeeld van een dergelijk instrument is het *assessment center* (AC). Bij het AC zijn vaak beoordelaars betrokken die een subjectief oordeel vellen over de capaciteiten, het gedrag of de aanleg van de kandidaat. Scoreverschillen tussen culturele groepen op het AC zijn kleiner dan op de intelligentietest. Desalniettemin blijven ze vrij substantieel (De Meijer et al., 2006). Omdat niet bekend is in welke mate subjectiviteit in beoordelingen bijdraagt aan deze scoreverschillen, onderzochten wij in deze studie wat de effecten zijn van gelijkens van beoordelaars ten aanzien van kandidaten in termen van culturele achtergrond.

We onderzochten de effecten van demografische en waargenomen gelijkens van beoordelaars ten aanzien van kandidaten op scores op een AC-opdracht, namelijk een rollenspel. Daarbij is onderscheid gemaakt tussen de autochtoon Nederlandse groep kandidaten, een Surinaams/voormalig-Antilliaanse groep en een Turkse/Marokkaanse groep. De data waren afkomstig van de Nederlandse Politieacademie en de selectie van hun politieagenten ($N_{\text{kandidaat}} = 27746$). Daarbij zijn hiërarchisch lineaire modellen (HLM) gebruikt om tegemoet te komen aan de geneste structuur

van onze data. Eerdere studies waarin eenzelfde methode van analyse werd gebruikt (McCarthy et al., 2010; Sacco et al., 2003), bestudeerden alleen demografische gelijkens en rapporteerden geen effecten van demografische gelijkens op interviewscores.

De vraag rijst vervolgens of ander gepubliceerd onderzoek naar demografische en waargenomen gelijkens-effecten, zónder inachtneming van de geneste structuur van data (zoals Buckley et al., 2007; McFarland et al., 2004; Turban et al., 2002), ten onrechte heeft geconcludeerd dat er significante effecten bestaan terwijl er in feite onvoldoende bewijs was voor het afwijzen van de *nul*-hypothese (Type-I-fout). De kans op een Type-I-fout wordt immers groter op het moment dat in de analyses geen rekening wordt gehouden met het feit dat er meerdere niveaus in de data aanwezig zijn, zoals het geval is in onze studie. Wij hebben dan ook meer vertrouwen gekregen in resultaten aan de hand van analyses die HLM hebben toegepast.

4.1 Bevindingen van deze studie

Onze resultaten met betrekking tot Hypothese 1 lieten zien dat demografische gelijkens, in dit geval culturele gelijkens, van de beoordelaar ten aanzien van de kandidaat resulteerde in een hogere AC-score op *Agency*. Dit effect werd niet gevonden voor de AC-score op *Communion*. Al met al werden er geen of slechts kleine effecten gevonden van demografische gelijkens op scores gemeten met het AC (Hypothese 1). Deze resultaten zijn in lijn met de resultaten van McCarthy et al. (2010) en Sacco et al. (2003) over beoordelingen gegeven op het selectie-interview. In ons onderzoek werd bovendien geen effect gevonden van waargenomen tussengroep-gelijkens op de scores gemeten met het AC (Hypothese 2). McCarthy et al.'s (2010) en Sacco et al.'s (2003) bevindingen onder een steekproef kandidaten in de VS over de niet-aanwezige effecten van demografische gelijkens op interviewscores kunnen dan ook worden gegeneraliseerd naar het AC, naar waargenomen tussengroep-gelijkens, en naar de Europese context.

Een eerste mogelijke verklaring voor het uitblijven van effecten van demografische en waargenomen gelijkens op de gegeven beoordelingen is dat beoordelaars tijdens personeelsselectie verantwoordelijk worden gehouden voor hun oordelen en dat zij daarom duidelijke redenen hebben om accuraat te zijn. De kosten die een verkeerd oordeel met zich meebrengen, zouden ervoor kunnen zorgen dat beoordelaars een grotere inspanning leveren in het beoordelingsproces dan mensen in het algemeen. In sociale interactie in het algemeen zal een waarnemer een globale indruk willen hebben van de persoon die waargenomen wordt. Als gevolg hiervan zal – in lijn met de sociale identiteitstheorie (Tajfel, 1982; Turner, 1987) en de zelf-categorisatietheorie

(Turner, 1987) – de waarnemer deze persoon categoriseren op basis van bepaalde kenmerken. Op het moment dat deze persoon geïdentificeerd wordt als lid van een bepaalde sociale groep (in ons geval, als lid van een culturele groep), zullen stereotypen die van toepassing zijn op deze sociale groep van invloed zijn – zowel direct als indirect – op de indruk die de waarnemer krijgt (Bodenhausen & Macrae, 1998). Hiertegenover stellen Kulik et al. (2007) dat sommige mensen egalitaire doelen hebben (d.w.z. doelen als eerlijkheid en gelijkheid in waarneming tussen verschillende groepen) en om die reden doen aan wat ook wel wordt aangeduid als ‘categorie-inhibitie’ (*category inhibition*). Deze categorie-inhibitie voorkomt dat een categorie invloed kan uitoefenen op sociale oordelen. Het is aannemelijk dat beoordelaars in situaties met grote belangen, zoals bij personeelsselectie, gericht zijn op gelijkheid en eerlijkheid tussen culturele groepen en daarom doen aan categorie-inhibitie, waardoor een gelijkenis of een verschil tussen beoordelaars en kandidaten geen effect zal hebben op de gegeven scores.

Een tweede mogelijke verklaring voor onze bevindingen is dat goedgetrainde beoordelaars geleerd hebben om zich te concentreren op een gestructureerde taak, waardoor zij minder beïnvloed zullen worden door aspecten als culturele gelijkenis of verschil. In lijn met deze gedachte toonden Hart et al. (2009) aan dat de motivatie om accuraat te zijn werd versterkt bij het maken van belangrijke beslissingen. Daarbij redeneerden Kulik et al. (2007) dat wanneer functie-eisen expliciet en objectief waren vastgesteld, kandidaten behorend tot een niet-positief geëvalueerde groep (in dit geval behorend tot een culturele minderheidsgroep) nog steeds positief geëvalueerd kunnen worden als zij voldeden aan de eisen van de functie. Ook De Kock en Hauptfleisch (2018) toonden recent aan dat het structureren van selectie-interviews samenhang met minder bias in beoordelingen.

Tot slot zou de algemene vraag gesteld kunnen worden of sociaalpsychologische theorieën met betrekking tot demografische en waargenomen gelijkenis altijd overeind blijven als zij uit het laboratorium gehaald worden en getest worden in de praktijk. Een derde verklaring voor het uitblijven van effecten van deze studie zou dan ook – in lijn met bevindingen die laten zien dat in het lab gevonden leeftijdsdiscriminatie tijdens selectie-interviews verdwijnt als studies buiten het lab worden uitgevoerd (Morgeson et al., 2008) – kunnen zijn dat effecten van demografische en waargenomen gelijkenis veel minder uitgesproken zijn in het echte leven. We willen er echter op wijzen dat zelfs kleine effecten van demografische of waargenomen gelijkenis, zoals we vonden in onze studie voor demografische gelijkenis, wel degelijk *adverse impact* kunnen veroorzaken tegenover minderheidsgroepen (Bartram, 1995). Dit kan voorkomen onder omstandigheden waarin streng wordt geselecteerd.

4.2 Beperkingen van deze studie

We hebben het onderzoek van McCarthy et al. (2010) en Sacco et al. (2003) over demografische gelijkenis uitgebreid door ook onderzoek te doen naar waargenomen gelijkenis. De gedachte hierachter is dat beoordelaars wellicht niet zozeer beïnvloed worden door feitelijke demografische gelijkenis, maar meer door waargenomen gelijkenis. Een eerste beperking van dit onderzoek is dat we waargenomen gelijkenis onderzochten door gebruik te maken van de percepties van beoordelaars over *groepen* kandidaten, ofwel de gemiddelde persoon in gedachten nemend uit een bepaalde culturele groep (d.w.z. Nederlands-Antilliaans, Marokkaans, Surinaams en Turks). Omdat we geen effect vonden op de beoordelingen op het AC, zou het kunnen zijn dat percepties van gelijkenis moeten worden gemeten op het individuele niveau (vgl. Landy & Farr, 1980), met andere woorden, aan de hand van de waargenomen gelijkenis ten aanzien van elke *individuele* autochtone Nederlandse kandidaat of kandidaat met een migratieachtergrond. Toekomstig onderzoek zou daarom waargenomen gelijkenis in meer detail moeten onderzoeken om te kunnen concluderen dat gelijkenis daadwerkelijk geen effect heeft op de gegeven beoordelingen.

Een tweede beperking van onze studie vormen de relatief kleine aantallen kandidaten met een migratieachtergrond ($n = 3089$) en beoordelaars met een migratieachtergrond ($n = 12$), tegenover de respectievelijk grote autochtoon Nederlandse groepen ($n = 24657$ kandidaten en $n = 135$ beoordelaars). Deze ongelijke groeps grootten resulteren in verlies aan onderscheidingsvermogen (*power*) en zouden mogelijk de effecten van demografische en waargenomen gelijkenis kunnen hebben verstoord.

Een derde beperking, die is gerelateerd aan het punt hierboven, is dat bij HLM *fixed* effecten lastig te detecteren zijn zonder een substantiële steekproef op Niveau-2 (Scherbaum & Ferrer, 2009). Specifiek voor deze studie zou dat betekenen dat een substantieel aantal beoordelaar/kandidaat-combinaties nodig is om effecten van demografische en waargenomen gelijkenis te kunnen opsporen. Vanwege het relatief kleine aantal beoordelaars met een migratieachtergrond zullen de verkregen resultaten naar alle waarschijnlijkheid een onderschatting zijn van de effecten. Toekomstig onderzoek zou daarom moeten blijven streven naar meer beoordelaars met een migratieachtergrond.

4.3 Conclusie

Dit onderzoek liet zien dat de beoordelingen op het AC niet of nauwelijks beïnvloed worden door demografische en waargenomen gelijkenis tussen cultureel diverse beoordelaars en cultureel diverse kandidaten. Daarom

werd er geen bewijs gevonden voor een differentieel effect van bias op de evaluaties van autochtoon Nederlandse kandidaten en kandidaten met een migratieachtergrond. Deze resultaten versterken onderzoeksbevindingen van McCarthy et al. (2010) en Sacco et al. (2003) naar demografische gelijkenis in de VS in relatie tot scores op het selectiegesprek, en breiden deze uit naar (1) het AC, (2) waargenomen gelijkenis, en (3) de Europese context. De studies van McCarthy et al. (2010), Sacco et al. (2003), en de huidige studie benadrukken het belang van het gebruik van correcte analysetechnieken voor geneste datastructuren.

Belangrijke verklaringen voor het uitblijven van effecten van demografische en waargenomen gelijkenis op gegeven beoordelingen kunnen zijn dat beoordelaars tijdens personeelsselectie gemotiveerd zijn om accurate beoordelingen te geven, en dat goedgetrainde beoordelaars geleerd hebben om zich te concentreren op een gestructureerde taak en daardoor minder beïnvloed zullen worden door aspecten als gelijkenis of verschil. Deze bevindingen verminderen de zorg in de praktijk dat discriminatie van kandidaten met een migratieachtergrond verklaard kan worden door gelijkenis of verschil tijdens professionele personeelsselectie. Toekomstig onderzoek moet echter blijven investeren in grotere steekproeven beoordelaars met een migratieachtergrond, zodat demografische en waargenomen gelijkenis meer gedetailleerd getoetst kunnen worden.

Praktijkbox

Wat betekenen de resultaten voor de praktijk?

- Bestaande verschillen in scores of beoordelingen op basis van selectie-instrumenten tussen groepen in de samenleving, zoals tussen culturele groepen, zullen te allen tijde kritisch moeten worden bekeken. Ogenscheinlijk kleine verschillen kunnen grote gevolgen hebben (*adverse impact*), bijvoorbeeld in geval van strengere selecties.
- Sociaalpsychologische theorieën blijven niet altijd overeen als zij uit het laboratorium gehaald worden en getest worden in de praktijk. Zo blijken de effecten van demografische en waargenomen gelijkenis minder uitgesproken in de praktijk van personeelsselectie dan verwacht werd op basis van theorie.
- Het assessment center (AC) is als selectie-instrument een goed alternatief voor veelgebruikte intelligentietests. Scoreverschillen op het AC tussen culturele groepen zijn kleiner dan op de intelligentietest, het AC is valide, en culturele bias van beoordelaars lijkt geen rol te spelen in de beoordeling van kandidaten.

Bijlage

Gemiddelde rang in de integratiehiërarchie zoals waargenomen door beoordelaars

Migratieachtergrond	Gemiddelde rang
Surinaams	3.57
Voormalig-Antilliaans	2.93
Turks	2.21
Marokkaans	1.29

NB. Hoe hoger de gemiddelde rang, hoe meer de groep wordt gezien als geïntegreerd in de Nederlandse samenleving.

Noot

1. Erkenning: We danken Hans van Loon van de Nederlandse Politieacademie voor zijn bijdrage aan dit paper.

Literatuur

- Bartram, D. (1995). Predicting adverse impact in selection testing. *International Journal of Selection and Assessment*, 3(1), 52-61.
- Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology: An International Review*, 46(1), 5-68.
- Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. *Advances in Social Cognition*, 11, 1-52.
- Bobko, P., & Roth, P. L. (2013). Reviewing, categorizing, and analyzing the literature on Black- White mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology*, 66(1), 91-126.
- Buckley, M. R., Jackson, K. A., Bolino, M. C., Veres, III, J. G., & Field, H. S. (2007). The influence of relational demography on panel interview ratings: A field experiment. *Personnel Psychology*, 60, 627-646.
- Byrne, D. (1971). *The attraction paradigm*. Academic Press.
- Carmichael, F., & Woods, R. (2000). Ethnic penalties in unemployment and occupational attainment: Evidence for Britain. *International Review of Applied Economics*, 14(1), 71-98.

- Centraal Bureau voor de Statistiek (CBS). (2016). *Bevolking: Kerncijfers*. <https://www.cbs.nl/nl-nl/achtergrond/2016/47/afbakening-generaties-met-migratie-achtergrond>
- Cohen, J. (1988). *Standard power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cox, T. H. (1993). *Cultural diversity in organizations: Theory, research and practice*. San Francisco, CA: Berrett-Koehler Publishers.
- Dayan, K., Fox, S., & Kasten, R. (2008). The preliminary employment interview as a predictor of assessment center outcomes. *International Journal of Selection and Assessment, 16*(2), 102-111.
- De Kock, F. S., & Hauptfleisch, D. B. (2018). Reducing racial similarity bias in interviews by increasing structure: A quasi-experiment using multilevel analysis. *International Perspectives in Psychology: Research, Practice, Consultation, 7*(3), 137-154.
- De Meijer, L. A. L., Born, M. Ph., Terlouw, G., & Van der Molen, H. T. (2006). Applicant and method factors related to ethnic score differences in personnel selection: A study at the Dutch police. *Human Performance, 19*(3), 219-251.
- Ensher, E. A., Grant-Vallone, E. J., & Marelich, W. D. (2002). Effects of perceived attitudinal and demographic similarity on protégés' support and satisfaction gained from their mentoring relationships. *Journal of Applied Social Psychology, 32*(7), 1407-1430.
- Ensher, E. A., & Murphy, S. E. (1997). Effects of race, gender, perceived similarity, and contact on mentor relationships. *Journal of Vocational Behavior, 50*(3), 460-481.
- Falk, A., & Fox, S. (2014). Gender and ethnic composition of assessment centers and its relationship to participants' success. *Journal of Personnel Psychology, 13*(1), 11-20.
- Ferris, G. R., & Judge, T. A. (1991). Personnel/human resources management: A political influence perspective. *Journal of Management, 17*(2), 447-488.
- Fiske, S. T., & Taylor, S. E. (2016). *Social cognition* (3rd ed.). New York: McGraw-Hill.
- Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology, 61*, 579-616.
- Fox, S., & Haboucha, S. (2011). Assessment centers in Israel: Some practices and perspectives. In G. Thornton & N. Povah (eds.), *Assessment centers and global talent management* (pp. 453-470). Gower.
- Gaugler, B. B., Rosenthal, D. B., Thornton III, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*(3), 493-511.
- Goldstein, H. W., Scherbaum, C. A., & Yusko, K. (2009). Adverse impact and measuring cognitive ability. In J. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes testing* (pp. 95-134). Psychology Press.

- Goldstein, H. W., Yusko, K. P., Braverman, E. P., Smith, D. B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology, 51*, 357-374.
- Guillaume, Y. R. F., Dawson, J. F., Otaye-Ebede, L., Woods, S. A., & West, M. A. (2017). Harnessing demographic differences in organizations: What moderates the effects of workplace diversity? *Journal of Organizational Behavior, 38*, 276-303.
- Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*(4), 555-588.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve*. New York: Free Press Paperbacks.
- Hough, L. M. (1998). Personality at work: Issues and evidence. In M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131-159). Erlbaum.
- Hox, J. (2010). *Multilevel analysis techniques and applications*. Lawrence Erlbaum Ass.
- Jennissen, R. (2009). Een algemeen beeld van internationale migratie. In Ministerie van Justitie (Red.), *Migratie naar en vanuit Nederland: een eerste proeve van de migratiekaart* (pp. 3-39). WODC.
- Klimoski, R. J., & Donahue, L. M. (2001). Person perception in organizations: An overview of the field. In M. London (Ed.), *How people evaluate others in organizations* (pp. 5-43). Lawrence Erlbaum.
- Kulik, C. T., Roberson, L., & Perry, E. L. (2007). The multiple-category problem: Category activation and inhibition in the hiring process. *Academy of Management Review, 32*(2), 529-548.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*(1), 72-107.
- Lankau, M. J., Riordan, Ch. M., & Thomas, Ch. H. (2005). The effects of similarity and liking in formal relationships between mentors and protégés. *Journal of Vocational Behavior, 67*(2), 252-265.
- McCarthy, J. M., Van Iddekinge, C. H., & Campion, M. A. (2010). Are highly structured job interviews resistant to demographic similarity effects? *Personnel Psychology, 63*, 325-359.
- McCroskey, J. C., Richmond, V. P., & Daly, J. A. (1975). The development of a measure of perceived homophily in interpersonal communication. *Human Communication Research, 1*, 323-332.
- McFarland, L. A., Ryan, A. M., Sacco, J. M., & Kriska, S. D. (2004). Examination of structured interview ratings across time: The effects of applicant race, rater race, and panel composition. *Journal of Management, 30*(4), 435-452.
- Modood, T. (1997). Employment. In T. Modood (Ed.), *Ethnic minorities in Britain* (pp. 83-149). Policy Studies Institute.

- Morgeson, F. P., Reider, M. H., Campion, M. A., & Bull, R. A. (2008). Review of research on age discrimination in the employment interview. *Journal of Business and Psychology, 22*(3), 223-232.
- Murphy, K. R. (2002). Can conflicting perspectives on the role of g in personnel selection be resolved? *Human Performance, 15*(1/2), 173-186.
- Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus user's guide* (6th ed.). Muthén & Muthén.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Sage.
- Sacco, J. M., Scheu, C. R., Ryan, A. M., & Schmitt, N. (2003). An investigation of race and gender similarity effects in interviews: A multilevel approach to relational demography. *Journal of Applied Psychology, 88*(5), 852-865.
- Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12*(2), 347-367.
- Segall, M. H., Dasen, P. R., Berry, J. W., & Poortinga, Y. H. (1999). *Human behavior in global perspective. An introduction to cross-cultural psychology* (2nd ed.). Allyn & Bacon.
- Strauss, J. P., Barrick, M. R., & Connerley, M. L. (2001). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and liking. *Journal of Occupational and Organizational Psychology, 74*(5), 637-657.
- Stupar, S., Van de Vijver, F., Te Lindert, A., & Fontaine, J. R. J. (2014). Multicultural attitudes mediate the relation between personality and perceived ethnic outgroup distance in the Netherlands. *International Journal of Intercultural Relations, 38*, 24-35.
- Tajfel, H. (1982). *Social identity and intergroup relations*. London: Cambridge University Press.
- Thornton, G. C. III, & Rupp, D. E. (2006). *Assessment centers in human resource management*. Erlbaum.
- Turban, D. B., Dougherty, T. W., & Lee, F. K. (2002). Gender, race, and perceived similarity effects in developmental relationships: The moderating role of relationship duration. *Journal of Vocational Behavior, 61*(2), 240-262.
- Turner, J. C. (1987). *Rediscovering the social group: A self-categorization theory*. Blackwell.
- Van den Berg, R. H., & Van Leest, P. (1999). Praktisch testgebruik bij allochtonen: Wanneer zinnig, nuttig of waardeloos? *Psycholoog, juni*, 256-260.
- Verkuyten, M., Hagendoorn, L., & Masson, K. (1996). The ethnic hierarchy among majority and minority youth in The Netherlands. *Journal of Applied Social Psychology, 26*(12), 1104-1118.

- Weijters, G., & Scheepers, P. (2003). Verschillen in sociale integratie tussen etnische groepen: Beschrijving en verklaring. *Mens en Maatschappij*, 78(2), 144-157.
- Wiggins, J. S., & Trapnell, P. D. (1996). A dyadic-interactional perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality* (pp. 88-162). The Guilford Press.

Correspondentieadres

Lonneke A.L. de Meijer, Erasmus Universiteit Rotterdam, Erasmus School of Social and Behavioral Sciences, Woudestein Mandeville-gebouw (T)16-27A, Postbus 1738, 3000 DR Rotterdam.

Alle materialen die in deze studie zijn gebruikt (d.w.z. data en modellen), kunnen bij de eerste auteur worden opgevraagd, demeijer@essb.eur.nl.