

SELECTIE BIJ MBO-STAGES

EEN EXPERIMENT MET FICTIEVE KANDIDATEN ONDER
RECRUITERS VAN MBO-STAGIAIRS

RAPPORT

seo • economisch onderzoek

AUTEURS

DEVI BRANDS, MARCEL COENDERS, HARRIËT PRINS, BAS TER WEEL

MET SUBSIDIE VAN



AMSTERDAM, SEPTEMBER 2023

Samenvatting

Uit een experiment onder recruiters blijkt dat stagekandidaten met een migratieachtergrond en dezelfde objectieve kenmerken slechter beoordeeld worden dan kandidaten zonder migratieachtergrond, vooral bij mannelijke Marokkaanse kandidaten. Een meer gestandaardiseerde en gestructureerde beoordelingsmethode verhelpt deze negatieve bias voor een deel.

Discriminatie in wervingsprocedures op de arbeidsmarkt is hardnekkig en fors, en met wetenschappelijk onderzoek in verschillende landen aangetoond. Er is nauwelijks onderzoek gedaan naar discriminatie bij de selectie van stagiairs. Dit onderzoek toetst in hoeverre er sprake is van een bias bij de werving van studenten met een migratieachtergrond in het middelbaar beroepsonderwijs (mbo) die solliciteren naar een stageplaats. Er is voor dit onderzoek een online experiment opgezet met een steekproef van 446 recruiters van geregistreerde leerbedrijven in Nederland die stageplaatsen aanbieden aan mbo-studenten. De recruiters beoordelen tijdens het experiment de cv's van (fictieve) sollicitanten op basis van een aantal objectieve criteria zoals studierichting en -niveau, werkervaring en meetbare persoonskenmerken, waarbij extra taalvaardigheden in de Arabische of Turkse taal een migratieachtergrond suggereren. De cv's van kandidaten worden op drie manieren beoordeeld: aan de hand van een cijfer tussen 1 en 10, op basis van een vraag of de kandidaat wordt uitgenodigd voor een gesprek en op basis van de vraag in hoeverre er sprake is van een match tussen de kandidaat en de organisatie die de recruiter vertegenwoordigt.

De eerste bevinding van het onderzoek is dat er sprake is van een statistisch significante negatieve bias ten opzichte van kandidaten met een migratieachtergrond als het gaat om de beslissing om een kandidaat uit te nodigen voor een gesprek en wanneer de match tussen de kandidaat en de organisatie moet worden beoordeeld. Er bestaat geen statistisch significant verschil tussen kandidaten in de beoordeling van hun cv met een cijfer van 1 tot en met 10. Wanneer de analyse wordt gesplitst naar de verschillende migratieachtergronden blijkt dat Marokkaanse kandidaten en in het bijzonder mannelijke Marokkaanse kandidaten een minder goede beoordeling krijgen. Deze resultaten zijn consistent met de wetenschappelijke literatuur die constateert dat negatieve biases vooral voorkomen bij het beoordelen van mannelijke leden van migrantengroepen. De bevindingen zijn relevant voor het mbo, omdat het lopen van een stage belangrijk is voor een succesvolle arbeidsmarktintrede. Het voorkomen van negatieve biases bij de werving van stagiairs draagt bij aan het verbeteren van kansengelijkheid in het onderwijs en op de arbeidsmarkt.

De tweede bevinding van het onderzoek is dat een meer gestandaardiseerde en gestructureerde methode voor het presenteren en beoordelen van cv's de negatieve bias vermindert. De conventionele methode, waarbij volledige cv's van kandidaten worden beoordeeld, wordt vergeleken met een methode waarbij eerst de verschillende onderdelen van elk cv afzonderlijk worden beoordeeld en die beoordelingen vervolgens worden meegenomen bij de beoordeling van het gehele cv. Uit de analyse komt naar voren dat de negatieve bias ten opzichte van kandidaten met een migratieachtergrond met betrekking tot een uitnodiging voor een gesprek statistisch significant kleiner is als de meer gestructureerde methode wordt toegepast. Voor de match met de organisatie treedt onder de gehele onderzoeksgroep geen statistisch significante vermindering van de negatieve bias op. Aanvullende analyses laten zien dat de gestandaardiseerde en gestructureerde methode succesvoller is in het verminderen van de negatieve bias bij de beoordeling van (mannelijke) Marokkaanse kandidaten: Marokkaanse

kandidaten (en in het bijzonder mannen) worden negatiever beoordeeld als het gaat om de waargenomen match tussen de kandidaat en de organisatie en bij de beslissing om een kandidaat uit te nodigen voor een gesprek, maar in beide gevallen is deze negatieve bias significant kleiner als de meer gestandaardiseerde en gestructureerde methode wordt gebruikt. Deze bevindingen wijzen erop dat vooral groepen die het meest geconfronteerd worden met een negatieve bias in een normale selectieprocedure baat zouden kunnen hebben bij een meer gestandaardiseerde en gestructureerde selectiemethode.

Leerbedrijven die stages aanbieden voor mbo-studenten hebben meestal een relatief kleine omvang. Zo werkt 70 procent van de recruiters uit ons onderzoek bij een leerbedrijf met minder dan 50 werknemers. Dergelijke kleine bedrijven hebben waarschijnlijk minder middelen dan een HR-afdeling van een grote onderneming om een objectieve selectieprocedure in te richten. De in dit experiment onderzochte gestandaardiseerde en gestructureerde methode om cv's te beoordelen is eenvoudig toe te passen, ook in kleinere bedrijven. Daarmee kan deze methode helpen bij een eerlijker en objectiever selectieproces, door een kleine en eenvoudige, en daarmee haalbare, verandering in de manier waarop cv's worden gepresenteerd en beoordeeld in te voeren.

Inhoudsopgave

Samenvatting		i
1	Inleiding	1
	1.1	Gestructureerde cv-beoordeling 1
	1.2	Het onderzoek in drie stappen 1
2	Probleem- en interventieanalyse	4
	2.1	Context 4
	2.2	Mechanismen en hypothesen 5
	2.3	Interventie 6
	2.4	Uitkomstmaten 9
	2.5	Onderzoeksmethode 9
3	Resultaten	10
	3.1	Beschrijvende statistieken 10
	3.2	Samenvatting van de resultaten 10
	3.3	Resultaten 13
	3.4	Resultaten voor subgroepen 16
	3.5	Robuustheid van de resultaten 19
4	Impact	20
Referenties		22
Bijlage A	Linear random intercept en OLS modellen	24
Bijlage B	Controle voor sociaalwenselijkheid	27

1 Inleiding

Dit onderzoek analyseert het effect van een meer gestandaardiseerde en gestructureerde cv-beoordeling op bias in het selectieproces van stagiairs. Het onderzoek maakt deel uit van een overkoepelend onderzoek gericht op de versterking van de positie van mbo'ers met een migratieachtergrond bij de overgang naar de arbeidsmarkt.

1.1 Gestructureerde cv-beoordeling

In de literatuur komen standaardisering en objectivering van sollicitatieprocedures naar voren als een mogelijke manier om discriminatie bij de werving en selectie te verminderen (Lippens et al., 2023). Dit onderzoek is gericht op het meten van een mogelijke negatieve bias bij de werving en selectie van stagiairs met een migratieachtergrond in het mbo in Nederland en op het toetsen in hoeverre deze bias kan worden verminderd door een meer gestandaardiseerd en gestructureerd proces voor het beoordelen van cv's te gebruiken. Het onderzoek is gericht op het mbo, omdat in dit type beroepsonderwijs een relatief grote groep jongeren met een migratieachtergrond studeert. Binnen het mbo zijn stages erg belangrijk voor het succesvol afronden van de opleiding en geeft de kwaliteit van de stage een signaal aan toekomstige werkgevers over de kwaliteit van de kandidaat. In een online experiment hebben recruiters van leerbedrijven de cv's van fictieve mbo-kandidaten beoordeeld en een oordeel gegeven over de mate waarin zij passen bij de organisatie en in hoeverre ze de sollicitant zouden uitnodigen voor een gesprek over het lopen van een stage. De beoordeling van cv's heeft op twee manieren plaatsgevonden: volgens de conventionele methode, waarbij een volledig cv beoordeeld is, en volgens de gestandaardiseerde en gestructureerde methode, waarbij eerst de afzonderlijke cv-onderdelen beoordeeld zijn en daarna pas het volledige cv.

1.2 Het onderzoek in drie stappen

Het onderzoek maakt deel uit van een overkoepelend onderzoek naar het identificeren van werkzame mechanismen bij het versterken van de positie van mbo'ers met een migratieachtergrond bij de overgang van het onderwijs naar de arbeidsmarkt. Het onderzoek is mogelijk gemaakt met subsidie van NRO. Binnen het onderzoek worden verschillende programma's geëvalueerd op hun effectiviteit in het versterken van de positie van deze groep mbo-studenten. Er wordt onderscheid gemaakt tussen programma's die gericht zijn op de fase voorafgaand aan het solliciteren en programma's die gericht zijn op de wervings- en selectiefase. Er zijn weinig interventies die betrekking hebben op de wervings- en selectiefase. Er is daarom een onderzoek uitgevoerd waarbij een interventie wordt gesimuleerd op basis van fictieve stagekandidaten.

Het onderzoek naar de gestandaardiseerde en gestructureerde cv-beoordeling is in drie stappen uitgevoerd. De eerste stap bevat de probleem- en interventieanalyse. Hierbij staat de volgende vraag centraal: wat is het probleem en hoe draagt dit onderzoek bij aan een oplossing? In stap 2 staan de resultaten van de verschillende beoordelingsmethoden centraal. Er wordt nagegaan in hoeverre sprake is van een negatieve bias bij de conventionele selectiemethode en in hoeverre en voor welke groepen deze bias afneemt door de gestandaardiseerde en gestructureerde methode toe te passen. Dit leidt ten slotte tot een conclusie als laatste stap

van het onderzoek. De conclusie is bedoeld om beleidsmakers en recruiters aanknopingspunten te geven om de mogelijke negatieve bias in de selectie van kandidaten te voorkomen.

1.2.1 Probleem- en interventieanalyse

De probleem- en interventieanalyse (stap 1 in de evaluatie) is uitgevoerd aan de hand van een zogenaamde CIMO-configuratie. Met behulp van literatuuronderzoek is de interventie opgesteld en zijn hypothesen gevormd over het mogelijke effect van de interventie. In de CIMO-configuratie is beschreven op welke manier de interventie hulp kan bieden aan stagezoekende mbo-studenten. Hierin staat centraal in welke context (C) de interventie (I) leidt tot outcome (O) en via welke veronderstelde mechanismen (M) de uitkomsten worden bereikt. Aanvullend is beschreven met welke onderzoeksmethoden de resultaten onderzocht zijn om de uitkomsten te kunnen interpreteren.

1.2.2 Resultaten

De tweede stap van de evaluatie beschrijft de resultaten. De resultaten zijn verkregen door een online experiment onder recruiters van geregistreerde leerbedrijven. De Samenwerkingsorganisatie Beroepsonderwijs Bedrijfsleven (SBB) voert namens het ministerie van Onderwijs, Cultuur en Wetenschap de accreditatie van deze leerbedrijven uit. Met hulp van SBB zijn geregistreerde leerbedrijven benaderd om een steekproef te verkrijgen van bedrijfseigenaren en HR-managers (hierna: recruiters) die ervaring hebben met het aannemen van mbo-studenten voor stages.

SBB heeft een steekproef getrokken van 8.000 leerbedrijven die in de afgelopen twee jaar mbo-stagiairs hebben aangenomen. De steekproef bestaat uit bedrijven van verschillende omvang, die actief zijn in verschillende sectoren en regio's. De recruiters is gevraagd deel te nemen aan een anoniem online onderzoek om bij te dragen aan het verbeteren van de arbeidsmarktkansen van mbo-studenten. Om het risico op sociaalwenselijke antwoorden te minimaliseren, is in de begeleidende uitnodiging niets vermeld over het doel om een mogelijke negatieve bias te meten bij het rekruteren van mbo-studenten voor stages.

Het onderzoek begint met een vraag of respondenten al dan niet betrokken zijn bij het aannemen van studenten. Bij respondenten die geen betrokkenheid hebben bij dit proces wordt de enquête beëindigd. In totaal hebben 446 respondenten de volledige enquête ingevuld: 46 procent van de respondenten is man, 52 procent is vrouw en 2 procent is non-binair of het geslacht is onbekend, waarbij de meeste respondenten (65 procent) tussen de 45 en 65 jaar oud zijn. 14 procent van de respondenten is buiten Nederland geboren of heeft ten minste één ouder voor wie dit het geval is. De meerderheid van de respondenten heeft een opleiding in het hoger beroepsonderwijs (hbo) afgerond (57 procent). Respondenten werken vooral bij bedrijven met minder dan 50 werknemers (70 procent). De middelgrote en grote bedrijven zijn evenredig verdeeld (15 procent uit bedrijven met 50-250 werknemers, 15 procent met meer dan 250 werknemers). De verdeling over regio's is representatief. Bedrijven zijn actief in verschillende economische sectoren en activiteiten, waarbij zorg, welzijn en sport de grootste sector vormt (33 procent).

1.2.3 Verklarende evaluatie/condities voor succes

Het onderzoek bestaat uit een fictieve toetsing van cv's door professionele recruiters. Dit heeft als voordeel dat de onderzoekers de setting waarin het onderzoek plaatsvindt nauwgezet kunnen bepalen om zo zuiver mogelijk een effectmeting uit te voeren. In dit geval betekent dat een onderscheid tussen de resultaten op het beoordelen, uitnodigen en matchen van stagekandidaten van de conventionele selectiemethode en de gestandaardiseerde en

gestructureerde selectiemethode. Een potentieel nadeel is dat de externe validiteit, conclusies en bredere toepasbaarheid van de resultaten in relatie tot het experiment mogelijk beperkt zijn. De resultaten van het onderzoek geven aanknopingspunten voor SBB en recruiters om het selectieproces anders in te richten, omdat blijkt dat een gestandaardiseerde en gestructureerde selectiemethode de negatieve bias voor jongeren met een migratieachtergrond verkleint. Het reduceren van de negatieve bias is zowel in het voordeel van de stagiair die een kwalitatief betere stage kan lopen als in het voordeel van het bedrijf waar het gaat om het opleiden van de toekomstige werknemers. Tevens blijkt dat deze gestandaardiseerde en gestructureerde selectiemethode niet de gehele negatieve bias wegneemt, wat vraagt om aanvullende analyses en mogelijke maatregelen om de negatieve bias in de toekomst verder te reduceren. Deze verdere reductie is ook in het belang van studenten en werkgevers, omdat er in de toekomst een beter opgeleide beroepsbevolking beschikbaar is en in het belang van de maatschappij als geheel omdat kansengelijkheid wordt gestimuleerd en arbeidspotentieel minder onbenut blijft. Ten slotte is de externe validiteit van de resultaten getoetst aan de hand van de wetenschappelijke literatuur, die tot vergelijkbare conclusies komt, en kan deze verder worden getoetst door de voorwaarden die worden gesteld aan het selectieproces bij een aantal bedrijven in te voeren en zo verder te onderzoeken in hoeverre dit in een duurzame vermindering van de negatieve bias resulteert.

2 Probleem- en interventieanalyse

Recruiters beoordelen zes fictieve cv's met de conventionele methode en zes met de gestandaardiseerde en gestructureerde methode. Bias wordt gemeten door de invloed van het cv-attribuut talenkennis (Turks of Marokkaans/Arabisch) op de beoordelingen. De hypothese is dat een gestandaardiseerde en gestructureerde selectiemethode bias vermindert.

2.1 Context

Er is steeds meer wetenschappelijke literatuur die het bestaan van wervingsdiscriminatie, oftewel discriminatie bij de werving en selectie van sollicitanten, aantoont. Veldexperimenten tonen bijvoorbeeld aan dat sollicitanten met een migratieachtergrond minder vaak worden teruggebeld door werkgevers dan even gekwalificeerde sollicitanten zonder migratieachtergrond (Lippens et al., 2023).

Er zijn verschillende verklaringen voor wervingsdiscriminatie. Twee klassieke verklaringen zijn 'smaak' of taste based-discriminatie en statistische discriminatie. Bij taste based-discriminatie komt wervingsdiscriminatie voort uit vooroordelen over sollicitanten met een migratieachtergrond (Becker, 1971). Werkgevers kunnen zelf bevooroordeeld zijn of rekening houden met dergelijke vooroordelen bij hun personeel of klanten. Wat betreft de oorsprong van deze vooroordelen, belicht psychologisch en sociologisch onderzoek de dynamiek van relaties tussen groepen. Negatieve stereotypen zijn gerelateerd aan een gebrek aan contact en begrip tussen verschillende groepen in de samenleving (Pettigrew & Tropp, 2011), gevoelens van dreiging (Scheepers et al., 2002), sociale dominantie-oriëntatie (Sidanius et al., 2004), en de sociale en culturele afstand tussen groepen (Verkuyten et al., 1996; Chambers et al., 2013). Er worden verschillende niveaus van discriminatie gevonden voor verschillende groepen, die overeenkomen met hun relatieve positie in de etnische hiërarchie, dat wil zeggen de rangschikking van voorkeuren voor verschillende etnische groepen in een samenleving (Hagendoorn, 1995). Dit komt overeen met de taste based-discriminatie-theorie. Zo wordt er meer discriminatie gevonden ten opzichte van groepen met een lage positie in de etnische hiërarchie, zoals sollicitanten met een islamitische, Afrikaanse of Midden-Oosterse achtergrond (Lancee, 2021; Lippens et al., 2023; Thijssen et al., 2022; Vernby & Dancygier, 2019), wat wordt toegeschreven aan een grotere sociale en culturele afstand tot de gemiddelde Nederlandse samenleving.

De tweede verklaring is statistische discriminatie. Bias is dan niet het resultaat van irrationele vooroordelen, maar van rationele acties van winstmaximaliserende actoren, die geconfronteerd worden met onzekerheid bij hun beslissingen in de werving en selectie van sollicitanten (Arrow, 1998; Phelps, 1972). Werkgevers hebben beperkte informatie over de productiviteit van sollicitanten en vertrouwen daarom op hun kennis of opvattingen over de gemiddelde productiviteit van een groep sollicitanten. Dit betekent dat zichtbare indicatoren over groepen (zoals etniciteit, geslacht, leeftijd) gebruikt worden om niet-waargenomen vaardigheden, competenties en gedragingen uit af te leiden. Werkgevers hebben een negatieve bias ten opzichte van sollicitanten met een migratieachtergrond als ze de groep als minder gekwalificeerd of betrouwbaar beschouwen. Een andere vorm van statistische discriminatie is dat werkgevers sollicitanten met een migratieachtergrond minder goed beoordelen als ze veronderstellen dat de variatie in de kwaliteit van kennis en vaardigheden groter is binnen de groep sollicitanten met een migratieachtergrond (Aigner & Cain, 1977). Dit zou betekenen dat werkgevers de gemiddelde productiviteit van sollicitanten zonder migratieachtergrond beter kunnen voorspellen dan de gemiddelde

productiviteit van sollicitanten met een migratieachtergrond door minder kennis van en bekendheid met de groep sollicitanten met een migratieachtergrond en vanwege culturele verschillen. Dit betekent ook dat de negatieve bias daalt als er meer objectieve informatie en kennis beschikbaar zijn over de kwaliteiten van sollicitanten. De reden daarvoor is dat de onzekerheid afneemt en men minder hoeft te vertrouwen op groepsstatistieken als ruwe, indirecte indicatoren voor de productiviteit (zie bijv. Kaas & Manger, 2012).

De laatste tijd krijgt de rol van de impliciete of onbewuste bias ten opzichte van minderheidsgroepen meer aandacht.¹ Er kan een onderscheid gemaakt worden tussen expliciete en impliciete associaties, overtuigingen en houdingen ten opzichte van verschillende groepen. Waar een expliciete bias en vooroordelen controleerbaar zijn en bewust tot uiting komen, worden een impliciete bias en vooroordeel onbewust geactiveerd (Nosek, 2007). Impliciete vooroordelen veroorzaken spontaan en moeilijk controleerbaar gedrag.² Lab- en veldexperimenten tonen aan dat wervingsdiscriminatie het gevolg is van zowel een expliciete als impliciete negatieve bias ten opzichte van groepen met een migratieachtergrond (Blommaert et al., 2012; Rooth, 2010). Daarnaast vinden Dovidio & Gaertner (2000) meer discriminatie wanneer de beslissingssituatie ambigu is. Zonder expliciete criteria hebben recruiters de neiging om eenvoudig beschikbare informatie te gebruiken, wat leidt tot een bias in de selectie van kandidaten. Volgens deze redenering zou het standaardiseren van selectieprocedures het effect van bias verminderen. Wolgast et al. (2017) vinden bij een onderzoek onder studenten dat er minder bias optreedt bij een gestructureerde selectietaak dan bij een ongestructureerde selectietaak. Gestructureerde en gestandaardiseerde sollicitatieformulieren stellen recruiters in staat om zich te richten op de relevante competenties van de sollicitanten – in plaats van op minder relevante persoonlijke eigenschappen – en hen op een objectievere manier te scoren, vooral in vergelijking met door sollicitanten zelf opgestelde cv's die uniformiteit ontberen (e.g., Derous & Ryan, 2019).

2.2 Mechanismen en hypothesen

Dit onderzoek draagt bij aan de literatuur door het effect te testen van een meer gestandaardiseerde en gestructureerde methode voor het presenteren en beoordelen van kandidaten bij een selectieprocedure. We kijken naar het effect op verschillende uitkomstmaten, namelijk: het cijfer dat recruiters geven aan de kandidaten, hun oordeel over de mate waarin de kandidaat bij de organisatie past en de keuze om de kandidaat uit te nodigen voor een gesprek. Het onderzoek is gericht op kandidaten met een Turkse of Marokkaanse achtergrond, die tot de grootste minderheidsgroepen in Nederland behoren en kunnen worden geconfronteerd met negatieve bias vanwege hun achtergrond. De verwachting is dat sollicitanten uit deze minderheidsgroepen minder positief worden beoordeeld dan sollicitanten zonder migratieachtergrond. Er wordt onderscheid gemaakt tussen mannelijke en vrouwelijke sollicitanten, omdat de verwachting is dat de kans op vooroordelen groter is bij mannelijke kandidaten uit een minderheidsgroep.³

¹ Dit onderzoek is gerelateerd aan de twee denkwijzen die Kahneman (2011) onderscheidt: 'systeem 1'-denken is snel, emotioneel, onbewust, automatisch piloot-denken en 'systeem 2'-denken is langzamer, bedachtzaam en logisch. Door 'systeem 1'-denken leiden impliciete vooroordelen tot discriminatie, zelfs op een onbewust niveau.

² Volgens het MODE-model (Fazio, 1990) kan de directe link tussen automatisch geactiveerde opvattingen (bijv. een impliciete bias ten opzichte van minderheidsgroepen) en het daaruit voortvloeiende gedrag (bijv. discriminatie) worden verbroken als een actor een sterke motivatie heeft om zich op een andere manier te gedragen en de mogelijkheid heeft om dat te doen (oftewel de tijd, energie en het vermogen om de invloed van automatisch geactiveerde biases te overwinnen). In lijn met deze redenering kunnen situaties die gepaard gaan met tijdsdruk, vermoeidheid en die stressvol zijn bevooroordeeld redeneren uitlokken (zie Bertrand en Duflo 2016 voor een overzicht).

³ Negatieve bias verschilt tussen mannelijke en vrouwelijke kandidaten. Volgens de outgroup male target hypothesis, afgeleid van de sociale dominantie-theorie (Sidanius en Pratto, 2001), is er meer negatieve bias ten opzichte van mannen

Twee manieren om cv's te presenteren en te beoordelen worden vergeleken. Gewoonlijk wordt het volledige cv van elke kandidaat in het geheel beoordeeld. Een algehele beoordeling kan echter gedeeltelijk gebaseerd zijn op (impliciete) vooroordelen omdat groepsattributen, zoals de naam of kennis van vreemde talen, gebruikt kunnen worden als snelle heuristieken. Het afzonderlijk beoordelen van de specifieke cv-onderdelen kan het risico op het gebruik van (impliciete) vooroordelen verkleinen en de focus op de relevante vaardigheden of competenties vergroten.

1. De eerste hypothese luidt als volgt: De verwachting is dat minder negatieve bias optreedt als de relevante vaardigheden en competenties van de kandidaten afzonderlijk en systematisch worden beoordeeld, in vergelijking met een beoordeling van het volledige cv in één keer.

De mate van bias kan variëren in de verschillende stappen in het selectieproces. Over het algemeen beoordelen recruiters in de eerste fase van het selectieproces eerst elke kandidaat en beslissen ze vervolgens welke kandidaten worden uitgenodigd voor een gesprek.

2. De tweede hypothese luidt als volgt: De verwachting is dat een sterkere negatieve bias optreedt bij de keuze voor het uitnodigen voor gesprekken, omdat dit een complexere beslissing is ten opzichte van een oordeel over de kwaliteit van het cv .

Ten slotte speelt de verwachte match tussen een kandidaat en het (leer)bedrijf een belangrijke rol in het selectieproces (Wolgast et al., 2018). Ook hier gaat het om een subjectieve inschatting die een complexe afweging vereist. Niettemin gebruiken recruiters hun inschatting als selectiecriteria.

3. De derde hypothese luidt als volgt: De verwachting is dat er een sterkere negatieve bias optreedt als recruiters de selectie van kandidaten baseren op een beoordeling van de mate waarin de sollicitant past bij het (leer)bedrijf ten opzichte van een oordeel over de kwaliteit van het cv

2.3 Interventie

Om de drie hypothesen te testen, is een vignetstudie onder recruiters uitgevoerd. Elk vignet bestaat uit een cv met zes attributen: leeftijd, geslacht, opleidingsniveau, werkervaring, stagebeoordeling en talenkennis. Deze zes attributen vormen samen een realistisch cv van een potentiële mbo-stagiair en zorgen voor voldoende variatie tussen de cv's. Elk attribuut heeft meerdere mogelijke niveaus (zie Tabel 2.1). Met deze attributen en niveaus zijn 936 verschillende vignetten mogelijk.⁴ Iedere recruiter krijgt twee sets van zes cv's toegewezen.

De kandidaten hebben een Nederlandse, Turkse of Marokkaanse achtergrond. Om de etnische afkomst aan te geven, gebruiken we het attribuut talenkennis. Kandidaten met een migratieachtergrond hebben een tweede moedertaal op hun cv: Turks of Arabisch (Marokkaans). Verder hebben alle sollicitanten Nederlands als moedertaal. Dit is een weerspiegeling van het feit dat het merendeel van de kandidaten met een Turkse of Marokkaanse

uit minderheidsgroepen dan vrouwen uit minderheidsgroepen (Navarette et al., 2010). Dit kan in het bijzonder gelden voor minderheden uit een islamitisch herkomstland, aangezien moslimmannen vaak worden afgeschilderd als agressief en gewelddadig (zie Sniderman en Hagendoorn, 2007). Daarom zou men een grotere negatieve bias verwachten ten opzichte van mannelijke sollicitanten met een Turkse of Marokkaanse migratieachtergrond dan tegen vrouwelijke sollicitanten met een dergelijke migratieachtergrond (cf. Vernby en Dancygier, 2019).

⁴ Er zijn 1.080 mogelijke combinaties. De beperking van het aantal stagebeoordelingen voor mbo-2 in acht nemend (zie de voetnoot in de tabel) zijn er 936 mogelijke vignetten.

migratieachtergrond in Nederland is geboren. Om de variatie in talenkennis te vergroten, bevatten de cv's ook Engelse taalvaardigheden.

Het hebben van een tweede moedertaal is onafhankelijk van de andere attributen op de cv's. Er is sprake van een negatieve bias als, gegeven alle andere attributen, kandidaten met Turks of Arabisch (Marokkaans) als moedertaal lagere beoordelingen krijgen.

Tabel 2.1 Overzicht van de vignetattributen en de verschillende niveaus

#	Attribuut	Niveaus
1	Leeftijd	1. Gemiddeld (afhankelijk van opleidingsniveau) 2. Relatief oud (gemiddeld + 2 jaar)
2	Geslacht	1. Man 2. Vrouw
3	Opleidingsniveau	1. Mbo-niveau 2 (BOL) 2. Mbo-niveau 3 (BOL) 3. Mbo-niveau 4 (BOL)
4	Werkervaring	1. Geen relevante werkervaring uit bijbaan 2. 1 jaar relevante werkervaring uit bijbaan 3. 2 jaar relevante werkervaring uit bijbaan
5	Stagebeoordeling ^a	1. Geen beoordeling vermeld op cv 2. Positieve beoordeling van 1 eerdere stage 3. Positieve beoordeling van 2 eerdere stages 4. Zeer positieve beoordeling uit 1 eerdere stage 5. Zeer positieve beoordeling uit 2 eerdere stages
6	Talenkennis ^b	1. Nederlands (moedertaal), Engels (gemiddeld) 2. Nederlands (moedertaal), Engels (goed) 3. Nederlands (moedertaal), Turks (moedertaal), Engels (gemiddeld) 4. Nederlands (moedertaal), Turks (moedertaal), Engels (goed) 5. Nederlands (moedertaal), Arabisch (Marokkaans, moedertaal), Engels (gemiddeld) 6. Nederlands (moedertaal), Arabisch (Marokkaans, moedertaal), Engels (goed)

Bron: SEO Economisch Onderzoek (2023). Noot: a Twee (zeer) positieve beoordelingen (niveau 3 en 5) werden niet toegewezen aan kandidaten met mbo-niveau 2, omdat de duur van de opleiding over het algemeen te kort is voor twee stages. b Niveaus 1 en 2 van het attribuut Talenkennis zijn relatief vaak toegewezen, om een realistischere verdeling van kandidaten te verkrijgen

Iedere recruiter beoordeelt twaalf cv's: zes via de conventionele methode en zes via de gestandaardiseerde en gestructureerde methode. De volgorde van de twee rondes wordt willekeurig toegewezen. In de ronde met de conventionele methode worden de zes cv's één voor één gepresenteerd, zoals geïllustreerd in Voorbeeld 1 in Figuur 2.1. Bij ieder cv wordt de recruiter gevraagd de kandidaat op drie manieren te beoordelen. Ten eerste geven de recruiters elke kandidaat een cijfer van 1 tot en met 10. Ten tweede beoordelen ze twee stellingen over de kandidaat: "Ik denk dat deze sollicitant goed binnen onze organisatie past" en "Ik zou deze sollicitant uitnodigen voor een gesprek", beide gemeten op een schaal van 1 (sterk mee oneens) tot 5 (sterk mee eens).

De gestandaardiseerde en gestructureerde beoordelingsmethode bestaat uit twee stappen. In de eerste stap van de beoordeling (zie Voorbeeld 2 in Figuur 2.1) wordt de recruiters gevraagd om elk attribuut (opleidingsniveau, werkervaring, stagebeoordeling en taalvaardigheid) afzonderlijk te beoordelen met een cijfer van 1 tot en met 10, voor alle zes de kandidaten. Vervolgens worden in de tweede stap de volledige cv's één voor één gepresenteerd, samen met de gegeven beoordeling van de attributen (zie Voorbeeld 2 in Figuur 2.1). De recruiters wordt nu gevraagd om elke kandidaat te beoordelen volgens dezelfde procedure als in de conventionele methode.

Figuur 2.1 Voorbeeld van beide rondes van de cv-beoordeling

Voorbeeld 1. Traditionele beoordelingsmethode:

Cv sollicitant 1	
Leeftijd	19 jaar
Geslacht	Man
Opleidingsniveau	Mbo-niveau 3 (BOL)
Werkervaring	Geen relevante werkervaring uit bijbaan
Stagebeoordeling	Positieve beoordeling van 1 eerdere stage
Talenkennis	Nederlands (moedertaal) Turks (moedertaal) Engels (goed)

Voorbeeld 2. Gestructureerde beoordelingsmethode

Stap 1. Het één voor één beoordelen van elk cv-attribuut (het voorbeeld is het attribuut 'Stagebeoordeling')

Sollicitant 1	Sollicitant 2	Sollicitant 3
Zeer positieve beoordeling uit 2 eerdere stages	Geen beoordeling vermeld op cv	Positieve beoordeling van 2 eerdere stages
<Score>	<Score>	<Score>

Sollicitant 4	Sollicitant 5	Sollicitant 6
Positieve beoordeling van 2 eerdere stages	Zeer positieve beoordeling van 1 eerdere stage	Positieve beoordeling van 1 eerdere stage
<Score>	<Score>	<Score>

Stap 2. Beoordeling van het volledige cv, met de eerder gegeven scores weergegeven naast het cv

Cv sollicitant 1	Uw beoordeling
Leeftijd	20 jaar
Geslacht	Man
Opleidingsniveau	Mbo-niveau 2 (BOL) 6
Werkervaring	1 jaar relevante werkervaring uit bijbaan 7

Stagebeoordeling	Zeer positieve beoordeling uit 2 eerdere stages	8
Talenkennis	Nederlands (moedertaal) Engels (gemiddeld)	7

2.4 Uitkomstmaten

Het onderzoek richt zich op drie uitkomstmaten.

- Ten eerste in hoeverre er verschillen zijn in de cv-beoordeling van studenten met een tweede moedertaal en studenten zonder tweede moedertaal;
- Ten tweede in hoeverre er een verschil is in de mate waarin recruiters aangeven dat deze kandidaten zouden passen bij hun organisatie;
- Ten derde in hoeverre er een verschil is in de mate waarin deze kandidaten worden uitgenodigd voor een gesprek.

Voor de drie uitkomstmaten wordt bepaald in hoeverre sprake is van een negatieve bias bij de conventionele methode en in hoeverre die verandert met de gestandaardiseerde en gestructureerde methode.

De drie uitkomstmaten leiden tot drie afhankelijke variabelen in de onderzoeksopzet: de beoordeling van de kandidaten (discreet, variërend van 1 tot en met 10) en de twee antwoorden op de stellingen over het passen bij de organisatie en de waarschijnlijkheid dat de kandidaat wordt uitgenodigd voor een gesprek (beide ordinaal, variërend van 1 [sterk mee oneens] tot 5 [sterk mee eens]). Aangezien de verdeling van de antwoorden op de twee stellingen erg scheef is, zijn beide variabelen omgezet in dummy's: één om aan te geven of de deelnemer het eens is met de stelling (1 indien (sterk) mee eens, anders 0) en één om aan te geven of de deelnemer het oneens is met de stelling (1 indien (sterk) mee oneens, anders 0).

2.5 Onderzoeksmethode

De data zijn hiërarchisch, aangezien elke respondent 12 vignetten beoordeelt. Een eenvoudige (*ordinary least squares* (OLS)) regressiemethode zou daarom vertekende resultaten opleveren. Om rekening te houden met de hiërarchische structuur en de dummy's voor de twee stellingen gebruiken we een *binomial generalized linear mixed model* met een *random intercept-model*.

De onafhankelijke variabelen voor ieder model dat wordt geschat zijn elk van de cv-attributen en een dummy die aangeeft of het cv al dan niet met de gestandaardiseerde en gestructureerde beoordelingsmethode is beoordeeld ('treatment dummy'). De cv-attributen worden opgenomen als dummyvariabelen voor elk van de niveaus. Omdat het effect van het hebben van een tweede moedertaal centraal staat, is ook de interactie tussen de treatment dummy en talenkennis meegenomen. De regressieanalyses zijn tot slot ook uitgevoerd met interacties tussen de treatment dummy, talenkennis en elk van de andere cv-attributen om te toetsen of er subgroepen kunnen worden geïdentificeerd waar het effect van een tweede moedertaal op de beoordeling sterker is.

3 Resultaten

Kandidaten met een tweede moedertaal worden minder vaak beschouwd als iemand die goed past bij de organisatie en minder vaak uitgenodigd voor een gesprek; vooral voor mannelijke Marokkaanse kandidaten. Een gestandaardiseerde en gestructureerde beoordelingsmethode vermindert deze negatieve bias deels.

3.1 Beschrijvende statistieken

Elke recruiter ($n=446$) heeft twaalf cv's beoordeeld, wat leidt tot een totaal van 5.352 beoordeelde cv's. De gemiddelde beoordeling is gelijk aan 7,0. Verder zijn recruiters het vaker eens met de stelling dat ze kandidaten zouden uitnodigen voor een gesprek dan met de stelling dat de kandidaten goed in hun organisatie zouden passen. Tabel 3.1 toont deze getallen, uitgesplitst naar de gemiddelden voor kandidaten met een tweede moedertaal en kandidaten zonder tweede moedertaal. Voor alle uitkomstmaten blijken kandidaten met een tweede moedertaal gemiddeld lager te worden beoordeeld dan kandidaten met alleen Nederlands als moedertaal. De verschillen in de gemiddelde beoordeling zijn kleiner bij de gestandaardiseerde en gestructureerde beoordelingsmethode dan bij de conventionele beoordelingsmethode. Voor beide beoordelingsmethoden zijn de gemiddelde verschillen klein.

Tabel 3.1 Sollicitanten met een tweede moedertaal worden gemiddeld iets lager beoordeeld

	Totaal			Conventionele methode			Gestructureerde methode		
Cv-beoordeling (1-10)									
Geen tweede moedertaal	7,03	1,32	2.735	7,03	1,38	1.370	7,03	1,26	1.365
Tweede moedertaal	7,00	1,40	2.617	6,97	1,44	1.306	7,03	1,36	1.311
Past bij de organisatie (1-5)									
Geen tweede moedertaal	3,58	0,87	2.735	3,59	0,87	1.370	3,57	0,88	1.365
Tweede moedertaal	3,54	0,88	2.617	3,54	0,87	1.306	3,55	0,89	1.311
Uitnodigen voor gesprek (1-5)									
Geen tweede moedertaal	3,83	0,93	2.735	3,85	0,92	1.370	3,81	0,93	1.365
Tweede moedertaal	3,79	0,95	2.617	3,79	0,95	1.306	3,79	0,96	1.311

Bron: SEO Economisch Onderzoek (2023)

3.2 Samenvatting van de resultaten

De resultaten van de regressiemodellen laten zien dat er geen statistisch significant verschil bestaat in de beoordeling van het cv (met een cijfer van 1 tot 10) van kandidaten met en zonder tweede moedertaal. Bij de stellingen over de match met de organisatie en het wel of niet uitnodigen voor een gesprek worden kandidaten met

een tweede moedertaal minder positief beoordeeld. De tweede en derde hypothese worden dus niet weerlegd: er is een sterkere negatieve bias ten opzichte van kandidaten met een migratieachtergrond bij het beoordelen van de match tussen kandidaat en organisatie en bij het wel of niet uitnodigen voor een gesprek, vergeleken met de beoordeling van de cv's met een cijfer.

Uitsplitsingen naar moedertaal en geslacht laten zien dat vooral Marokkaanse kandidaten, en met name Marokkaanse mannelijke kandidaten, minder goed worden beoordeeld dan kandidaten zonder tweede moedertaal.

Bij het gebruik van de meer gestandaardiseerde en gestructureerde methode wordt het verschil in de beoordeling van kandidaten met en zonder tweede moedertaal slechts ten dele kleiner. De negatieve bias ten aanzien van kandidaten met een tweede moedertaal in de kans om niet uitgenodigd te worden voor een gesprek verdwijnt grotendeels als gebruik wordt gemaakt van de meer gestandaardiseerde en gestructureerde methode om cv's te presenteren en te beoordelen. Maar bij de andere uitkomstmaten vinden we geen significant effect van deze methode.

Als we daarentegen kijken naar de effecten onder specifieke subgroepen van kandidaten (uitgesplitst naar moedertaal en geslacht), dan zien we in bepaalde subgroepen meer substantiële effecten van de meer gestandaardiseerde en gestructureerde methode. Zo vinden we een significante afname van de negatieve bias voor Marokkaanse kandidaten (in vergelijking tot kandidaten zonder migratieachtergrond) bij het gebruik van de gestandaardiseerde en gestructureerde methode, zowel voor het oordeel dat de kandidaat niet past bij de organisatie als voor de beslissing om de kandidaat niet uit te nodigen. Voor de subgroep van mannelijke Marokkaanse kandidaten vinden we dezelfde effecten en bovendien gaat de meer gestandaardiseerde en gestructureerde methode hierbij ook gepaard met significant minder negatieve bias bij de beslissing om de kandidaat wel uit te nodigen. Kortom, in het bijzonder bij die subgroepen waarbij de negatieve bias groot is – namelijk (mannelijke) Marokkaanse kandidaten – laten de resultaten meer significante effecten zien van de gestandaardiseerde en gestructureerde methode. Deze bevindingen zijn dus deels in overeenstemming met de verwachting (hypothese 1) dat er minder negatieve bias is als de relevante vaardigheden en competenties van de kandidaten afzonderlijk en systematisch worden beoordeeld, in vergelijking met een beoordeling van het volledige cv in één keer.

3.2.1 Resultaten voor de 'gemiddelde' sollicitant

De hierboven genoemde resultaten worden uitgebreid gepresenteerd in paragraaf 3. In de aldaar gepresenteerde regressiemodellen voor de antwoorden op de stellingen (Past de kandidaat bij de organisatie? Zou u de kandidaat uitnodigen voor een gesprek?) is de afhankelijke variabele steeds een dummy. Daarom is daar een *binomial random intercept model* gebruikt. Dit betekent echter dat de coëfficiënten geïnterpreteerd moeten worden als *odds ratios*, waardoor het niet eenvoudig is om snel te zien wat de omvang van het effect is. Om die reden bespreken we hier een aanvullende presentatie van de resultaten, door na te gaan wat dit betekent voor een 'gemiddelde' sollicitant. Op basis van het model hebben we voor de 'gemiddelde sollicitant' berekend wat de kans is dat de gemiddelde sollicitant wel of niet wordt gezien als match met de organisatie en wat de kans is dat de gemiddelde sollicitant wel of niet wordt uitgenodigd voor een gesprek. Deze resultaten, uitgesplitst naar moedertaal en beoordelingsmethode, zijn voor de 'gemiddelde sollicitant' samengevat in Tabel 3.2.

Tabel 3.2 is geordend in vier panels. In de eerste kolom wordt aangegeven in hoeverre een kandidaat niet of wel past bij de organisatie waarvoor de recruiter werkt. In de tweede kolom wordt aangegeven in hoeverre een

kandidaat niet of wel wordt uitgenodigd voor een gesprek. Daarnaast wordt binnen iedere kolom onderscheid gemaakt tussen de conventionele en gestructureerde methode om een cv te beoordelen.

De berekeningen laten in het algemeen zien dat de verschillen voor de ‘gemiddelde sollicitant’ relatief klein zijn tussen beide methodes. De gemiddelde Marokkaanse mannelijke kandidaat wordt minder vaak gezien als goede match met de organisatie en wordt minder vaak uitgenodigd voor een gesprek. Voor gemiddelde Marokkaanse mannelijke kandidaten is de kans om niet te worden gezien als match met de organisatie 0,25 procent en voor gemiddelde mannelijke kandidaten zonder tweede moedertaal 0,07 procent. De kans om niet te worden uitgenodigd voor een gesprek is voor gemiddelde Marokkaanse mannelijke kandidaten 0,31 procent en voor gemiddelde mannelijke kandidaten zonder tweede moedertaal 0,07 procent.

Met de gestructureerde methode verdwijnen deze verschillen grotendeels. Zo is bij de gestructureerde methode de kans dat een gemiddelde Marokkaanse mannelijke kandidaat niet wordt gezien als match met de organisatie 0,06 procent en voor gemiddelde mannelijke kandidaten zonder tweede moedertaal 0,07 procent. De kans om niet te worden uitgenodigd voor een gesprek is bij de gestructureerde methode voor de gemiddelde Marokkaanse mannelijke kandidaat 0,05 procent en voor de gemiddelde mannelijke kandidaat zonder tweede moedertaal 0,07 procent. Het verschil ten opzichte van de kans bij de conventionele methode is statistisch significant.

Ook bij de kans dat een recruiter het eens is met de stellingen zijn statistisch significante verschillen tussen de gemiddelde Marokkaanse mannelijke kandidaat en de gemiddelde mannelijke kandidaat zonder tweede moedertaal. De berekende kans dat de gemiddelde Marokkaanse mannelijke kandidaat wordt gezien als goede match met de organisatie is 62 procent en voor de gemiddelde mannelijke kandidaat zonder tweede moedertaal is dat 70 procent. De kans dat de gemiddelde Marokkaanse mannelijke kandidaat wordt uitgenodigd voor een gesprek is 93 procent en voor de gemiddelde mannelijke kandidaat zonder tweede moedertaal is dat 96 procent.

Bij de gestructureerde methode zijn de verschillen tussen de kandidaten kleiner, maar het verschil tussen de kans met de conventionele methode en de gestructureerde methode is niet altijd statistisch significant. De kans dat een gemiddelde Marokkaanse mannelijke kandidaat wordt uitgenodigd voor een gesprek is wel hoger met gebruik van de gestructureerde methode dan met de conventionele methode. De kans is dan 96 procent voor de gemiddelde Marokkaanse mannelijke kandidaat en eveneens 96 procent voor mannelijke kandidaten zonder tweede moedertaal.

Tabel 3.2 Berekende kans voor de gemiddelde kandidaat dat een recruiter het (niet) eens is met de stelling

	Deze kandidaat zou niet bij mijn organisatie passen		Ik zou deze kandidaat niet uitnodigen voor een gesprek	
	Conventionele methode	Gestructureerde methode	Conventionele methode	Gestructureerde methode
Tweede moedertaal				
Geen tweede moedertaal	0,04%	0,04%	0,08%	0,08%
Tweede moedertaal	0,07% ^{aa}	0,05%	0,14% ^{aa}	0,09% ^b
Specifieke tweede moedertaal				
Geen tweede moedertaal	0,06%	0,06%	0,08%	0,08%
Arabisch (Marokkaans)	0,11% ^{aa}	0,06% ^{bb}	0,18% ^{aa}	0,08% ^{bb}
Turks	0,09% ^a	0,07%	0,12%	0,11%
Specifieke tweede moedertaal, mannen				

Geen tweede moedertaal	0,07%	0,07%	0,07%	0,07%
Arabisch (Marokkaans)	0,25% ^{aa}	0,06% ^{bb}	0,31% ^{aa}	0,05% ^{bb}
Turks	0,11%	0,07%	0,10%	0,10%
Specifieke tweede moedertaal, vrouwen				
Geen tweede moedertaal	0,08%	0,08%	0,11%	0,11%
Arabisch (Marokkaans)	0,11%	0,08%	0,16%	0,11%
Turks	0,14%	0,10%	0,15%	0,15%
	Deze kandidaat zou bij mijn organisatie passen		Ik zou deze kandidaat uitnodigen voor een gesprek	
	Conventionele methode	Gestructureerde methode	Conventionele methode	Gestructureerde methode
Tweede moedertaal				
Geen tweede moedertaal	70,8%	70,8%	95,3%	95,3%
Tweede moedertaal	66,2% ^a	63,5%	94,2% ^a	93,6%
Specifieke tweede moedertaal				
Geen tweede moedertaal	71,7%	71,7%	95,2%	95,2%
Arabisch (Marokkaans)	64,6% ^{aa}	65,9%	93,0% ^{aa}	94,1%
Turks	69,1%	62,8%	95,0%	92,4% ^b
Specifieke tweede moedertaal, mannen				
Geen tweede moedertaal	70,0%	70,0%	95,5%	95,5%
Arabisch (Marokkaans)	61,7% ^a	63,5%	92,7% ^{aa}	95,8% ^b
Turks	69,7%	59,0%	95,0%	92,8%
Specifieke tweede moedertaal, vrouwen				
Geen tweede moedertaal	73,9%	73,9%	94,9%	94,9%
Arabisch (Marokkaans)	68,5%	69,5%	94,2%	91,4%
Turks	69,4%	66,2%	94,9%	92,1%

^a Significant verschil ($p < 0,10$) ten opzichte van 'Geen tweede moedertaal'

^{aa} Significant verschil ($p < 0,05$) ten opzichte van 'Geen tweede moedertaal'

^b Significant kleiner verschil ($p < 0,10$) ten opzichte van kandidaten zonder tweede moedertaal in de gestructureerde methode dan in de conventionele methode.

^{bb} Significant kleiner verschil ($p < 0,05$) ten opzichte van kandidaten zonder tweede moedertaal in de gestructureerde methode dan in de conventionele methode.

Bron: SEO Economisch Onderzoek (2023). De gemiddelde sollicitant is hier gedefinieerd als iemand met opleidingsniveau mbo-3, één jaar relevante werkervaring en één positieve stagebeoordeling uit een eerdere stage.

3.3 Resultaten

Voor elke uitkomstmaat is een regressiemodel gebruikt om het effect van een tweede moedertaal op de beoordeling te toetsen, waarbij gecontroleerd wordt voor de andere cv-kenmerken. Dit leidt tot vijf verschillende modellen: een model met de beoordeling met een cijfer als afhankelijke variabele en vier modellen met de dummy's van de stellingen als afhankelijke variabele (voor beide stellingen een dummy voor oneens en een dummy voor eens). Het effect van het hebben van een tweede moedertaal geeft aan of er sprake is van een bias bij de beoordeling. De interactie van de treatment dummy met het hebben van een tweede moedertaal geeft aan wat het extra effect is van de gestandaardiseerde en gestructureerde beoordelingsmethode op de beoordeling van

kandidaten met een tweede moedertaal ten opzichte van kandidaten zonder tweede moedertaal. Als de gestandaardiseerde en gestructureerde beoordelingsmethode de beoordeling van kandidaten met een tweede moedertaal sterker beïnvloedt dan de beoordeling van kandidaten zonder tweede moedertaal (en in tegengestelde richting), wijst dit erop dat de gestandaardiseerde en gestructureerde beoordelingsmethode bias vermindert.

De geschatte coëfficiënten zijn weergegeven in Tabel 3.3. In alle modellen worden kandidaten met een hoger opleidingsniveau, meer werkervaring en positieve beoordelingen van eerdere stages beter beoordeeld. Dit betekent dat ze significant vaker een hogere beoordeling krijgen, vaker worden beschouwd als een goede match met de organisatie en vaker worden uitgenodigd voor een gesprek. Leeftijd, geslacht en Engelse taalvaardigheden lijken in deze setting niet uit te maken, met uitzondering van vrouwen die gemiddeld een iets hoger cijfer ontvangen van de recruiters.

3.3.1 Effect tweede moedertaal op beoordeling

Bij de beoordeling van cv's met een cijfer heeft het hebben van een tweede moedertaal geen significant effect. We vinden dus geen negatieve bias bij de beoordeling van het cv met een cijfer op een schaal van 1 tot en met 10. In de modellen met de dummy's voor de stellingen vinden we wel statistisch significante effecten van het hebben van een tweede moedertaal. Zowel de kans om uitgenodigd te worden voor een gesprek als de kans op een positief oordeel over de match met de organisatie is significant lager voor kandidaten met een tweede moedertaal dan voor kandidaten zonder tweede moedertaal.

De geschatte coëfficiënten kunnen geïnterpreteerd worden als *odds ratios*. Dat betekent dat de geschatte kans dat recruiters een kandidaat niet bij de organisatie vinden passen 1,67 keer zo hoog is voor kandidaten met een tweede moedertaal en dat de geschatte kans dat recruiters niet bereid zijn om een kandidaat uit te nodigen voor een gesprek 1,66 keer zo hoog is voor kandidaten met een tweede moedertaal. Ook vinden recruiters het 0,80 keer zo waarschijnlijk dat kandidaten met een tweede moedertaal bij de organisatie zouden passen en 0,79 keer zo waarschijnlijk dat ze een kandidaat met een tweede moedertaal zouden uitnodigen voor een gesprek dan voor kandidaten zonder tweede moedertaal.

Deze *odds ratios* zijn een goede indicatie van de negatieve bias in de beoordeling van kandidaten met een tweede moedertaal relatief ten opzichte van de kandidaten zonder tweede moedertaal. Maar deze odds ratios kunnen een vertekend beeld geven als de werkelijke absolute kans op een goede of slechte beoordeling erg hoog of erg laag is. Daarom hebben we met het model ook de kans voor een 'gemiddelde kandidaat' berekend, zoals weergegeven in Tabel 3.2. Op basis van die berekeningen is de kans dat een recruiter een kandidaat niet wil uitnodigen voor een gesprek 0,14 procent voor een kandidaat met een tweede moedertaal en 0,08 procent voor een kandidaat zonder tweede moedertaal. Hieruit blijkt dat, hoewel het verschil relatief groot is, de kans voor beide groepen laag is. De kans dat een recruiter bereid is een kandidaat uit te nodigen voor een gesprek is 94,2 procent voor kandidaten met een tweede moedertaal en 95,4 procent voor kandidaten zonder tweede moedertaal. De kans dat een recruiter een kandidaat vindt passen bij de organisatie is 66,2 procent voor kandidaten met een tweede moedertaal en 70,8 procent voor kandidaten zonder tweede moedertaal.

Voor een eenvoudige interpretatie van de coëfficiënten hebben we ook de resultaten van een *linear random-intercept model* opgenomen in Bijlage A. Deze coëfficiënten kunnen worden geïnterpreteerd als verschillen in procentpunten. De lineaire regressieresultaten laten zien dat het verschil tussen kandidaten met en zonder een tweede moedertaal ruwweg twee procentpunt is, in alle modellen (Tabel A.1).

Tabel 3.3 Kandidaten met een tweede moedertaal worden minder vaak gezien als iemand die goed past bij de organisatie en minder vaak uitgenodigd voor een gesprek

	Beoordeling	Match met organisatie		Uitnodigen voor een gesprek	
		Past niet goed	Past goed	Nee	Ja
Talenkennis (naast Engels) (geen tweede moedertaal als referentiecategorie)					
Tweede moedertaal (Turks of Arabisch (Marokkaans))	-0,033 (0,033)	0,515** (0,204)	-0,218* (0,127)	0,506*** (0,192)	-0,241* (0,135)
Tweede moedertaal * Treatment	-0,013 (0,048)	-0,365 (0,282)	-0,110 (0,180)	-0,457* (0,267)	-0,090 (0,190)
Leeftijd (gemiddeld als referentiecategorie)					
Relatief oud	0,002 (0,024)	0,039 (0,138)	0,019 (0,088)	0,115 (0,131)	-0,016 (0,093)
Geslacht (man als referentiecategorie)					
Vrouw	0,061** (0,024)	0,040 (0,138)	0,067 (0,088)	-0,030 (0,131)	0,061 (0,093)
Opleidingsniveau (niveau 2 als referentiecategorie)					
Mbo-niveau 3	0,455*** (0,035)	-1,724*** (0,208)	1,234*** (0,128)	-1,514*** (0,193)	1,212*** (0,134)
Mbo-niveau 4	0,669*** (0,034)	-2,396*** (0,223)	1,779*** (0,131)	-2,075*** (0,202)	1,743*** (0,138)
Werkervaring (geen relevante werkervaring als referentiecategorie)					
1 jaar	0,349*** (0,029)	-0,734*** (0,169)	1,043*** (0,110)	-0,676*** (0,160)	1,015*** (0,116)
2 jaar	0,497*** (0,029)	-0,963*** (0,171)	1,407*** (0,112)	-0,860*** (0,161)	1,254*** (0,117)
Stagebeoordeling (geen beoordeling vermeld als referentiecategorie)					
1 positieve beoordeling	0,384*** (0,038)	-0,769*** (0,195)	1,097*** (0,138)	-0,912*** (0,187)	1,086*** (0,139)
2 positieve beoordelingen	0,566*** (0,042)	-1,180*** (0,262)	1,825*** (0,159)	-1,238*** (0,242)	1,565*** (0,168)
1 zeer positieve beoordeling	0,538*** (0,038)	-1,198*** (0,209)	1,610*** (0,143)	-1,275*** (0,198)	1,404*** (0,145)
2 zeer positieve beoordelingen	0,629*** (0,041)	-1,134*** (0,262)	1,881*** (0,159)	-1,517*** (0,254)	1,721*** (0,138)
Talenkennis Engels (gemiddeld als referentiecategorie)					
Goed	0,038 (0,024)	-0,015 (0,139)	0,079 (0,089)	-0,031 (0,132)	0,041 (0,094)
Treatment	-0,038 (0,033)	0,320 (0,196)	0,075 (0,124)	0,368** (0,186)	-0,112 (0,132)
Constante	5,880*** (0,064)	-4,520*** (0,636)	-1,411*** (0,233)	-4,009*** (0,604)	-0,293 (0,227)

Bron: SEO Economisch Onderzoek (2023). n vignetten = 5,352, n respondenten = 446, *p<0,10, **p<0,05, ***p<0,01

3.3.2 Effect gestructureerde beoordelingsmethode

We vinden dus een negatieve bias ten opzichte van kandidaten met een tweede moedertaal bij de conventionele beoordelingsmethode. Om te onderzoeken in hoeverre de gestandaardiseerde en gestructureerde beoordelingsmethode deze bias vermindert, is de interactieterm tussen de tweede moedertaal en de treatment dummy relevant. Dit geeft het extra effect van de gestructureerde beoordelingsmethode aan voor kandidaten met een tweede moedertaal. Dit betreft de geschatte coëfficiënten in de tweede rij van Tabel 3.3. Als deze coëfficiënt even groot is als de coëfficiënt in de eerste rij, maar met een omgekeerd teken, duidt dit op het verdwijnen van de bias door de gestructureerde beoordelingsmethode.

Voor de beoordeling van de cv's (de eerste kolom in Tabel 3.3) is geen significant verschil gevonden in de beoordeling van kandidaten met en zonder tweede moedertaal. Het toepassen van de gestandaardiseerde en gestructureerde beoordelingsmethode heeft ook geen significant verschillend effect op kandidaten met en zonder tweede moedertaal. Voor de beoordeling van een cv op een schaal van 1 tot en met 10 maakt het in deze setting en met deze recruiters dus niet uit hoe de beoordeling van kandidaten plaatsvindt.

Bij de stellingen over het passen bij de organisatie en het uitnodigen voor een gesprek is wel een significant lagere beoordeling van kandidaten met een tweede moedertaal gevonden in de conventionele methode. De coëfficiënten van de interactieterm zijn echter slechts in één van de modellen significant met een significantieniveau van 10 procent, namelijk het model dat de kans op het niet uitnodigen voor een gesprek schat. Oftewel, wat betreft de mate waarin de recruiter beoordeelt dat de kandidaat past (of niet past) bij de organisatie, verschilt het effect van de gestandaardiseerde en gestructureerde beoordelingsmethode niet tussen kandidaten met en zonder tweede moedertaal, omdat beide interactiecoëfficiënten niet significant van nul verschillen. Dit suggereert dat de negatieve bias ten opzichte van kandidaten met een tweede moedertaal blijft bestaan.

Hetzelfde geldt voor de bereidheid om een kandidaat uit te nodigen voor een gesprek. Ook hier vinden we geen significant verschillend effect van de gestandaardiseerde en gestructureerde beoordelingsmethode tussen kandidaten met en zonder tweede moedertaal. Wat betreft de kans om niet uitgenodigd te worden voor een gesprek, zien we dat de negatieve bias grotendeels teniet wordt gedaan door de gestandaardiseerde en gestructureerde beoordelingsmethode. De eerder berekende kansen laten zien dat dit voor de gemiddelde kandidaat betekent dat de kans dat een kandidaat met een tweede moedertaal niet wordt uitgenodigd voor een gesprek daalt van 0,14 procent (conventionele methode) naar 0,09 procent, waar dit voor de gemiddelde kandidaat zonder tweede moedertaal gelijk blijft op 0,08 procent. De resultaten laten zien dat de verschillen tussen kandidaten met en zonder tweede moedertaal in de kans om niet te worden uitgenodigd voor een gesprek significant kleiner zijn als de cv's met de gestructureerde beoordelingsmethode worden beoordeeld dan met de conventionele methode.

3.4 Resultaten voor subgroepen

De gemiddelde effecten zijn qua omvang niet bijzonder groot, maar voor subgroepen worden de effecten substantiëler. Er zijn modellen geschat die rekening houden met een mogelijke bias voor specifieke groepen en voor groepsspecifieke effecten van de treatment. Dat zijn dezelfde modellen als hierboven plus het onderscheid tussen verschillende moedertalen (Turks en Marokkaans-Arabisch) en geslacht. Alle andere cv-attributen zijn opnieuw opgenomen in de modellen. De geschatte coëfficiënten zijn samengevat in Tabel 3.4. De belangrijkste resultaten uit Tabel 3.3 hebben we ter vergelijking opnieuw opgenomen in de eerste twee rijen van Tabel 3.4.

3.4.1 Verschillen in beoordeling voor subgroepen

Door onderscheid te maken tussen de verschillende tweede moedertalen is te zien dat Arabisch (Marokkaans) als tweede moedertaal een significant negatief effect heeft op de beoordeling van de twee stellingen, terwijl dit voor Turks als tweede moedertaal alleen het geval is bij de kans om niet te worden gezien als goede match met de organisatie. Dit suggereert dat de eerder gevonden resultaten voornamelijk gedreven worden door een negatieve bias ten opzichte van Marokkaanse kandidaten. Door vervolgens ook onderscheid te maken tussen mannen en vrouwen blijkt dat er vooral een negatieve bias zichtbaar is ten opzichte van mannelijke kandidaten met Marokkaans-Arabisch als tweede moedertaal.

Voor een betere interpretatie van de coëfficiënten zijn opnieuw kansen berekend voor een gemiddelde kandidaat. De kans dat een recruiter een kandidaat niet wil uitnodigen voor een gesprek is 0,31 procent voor een mannelijke kandidaat met Marokkaans-Arabisch als tweede moedertaal en 0,07 procent voor hun medestudenten zonder tweede moedertaal. Hieruit blijkt opnieuw dat hoewel het verschil relatief groot is, de kans om niet uitgenodigd te worden voor beide groepen laag is. De kans dat een recruiter bereid is een kandidaat uit te nodigen voor een gesprek is 92,7 procent voor mannelijke kandidaten met Marokkaans-Arabisch als tweede moedertaal en 95,5 procent voor kandidaten zonder tweede moedertaal. De kans dat een recruiter de kandidaat niet vindt passen bij de organisatie is 0,25 procent voor een mannelijke kandidaat met Marokkaans-Arabisch als tweede moedertaal en 0,07 procent voor een kandidaat zonder tweede moedertaal. De kans dat een recruiter een kandidaat wel vindt passen bij de organisatie is 61,7 procent voor mannelijke kandidaten met Marokkaans-Arabisch als tweede moedertaal en 70,0 procent voor kandidaten zonder tweede moedertaal.

De lineaire coëfficiënten in de bijlage laten zien dat het verschil tussen kandidaten met en zonder Marokkaans-Arabisch als tweede moedertaal ruwweg drie procentpunt is in alle modellen. Dit verschil neemt toe bij mannelijke kandidaten (Bijlage A, Tabel A.2). Onze resultaten laten een vier procentpunt hogere kans zien dat mannelijke kandidaten met Marokkaans-Arabisch als tweede moedertaal niet als goede match met de organisatie worden beschouwd en een 3,6 procentpunt lagere kans dat zij als goede match met de organisatie worden beschouwd. Wat betreft de kans om uitgenodigd te worden voor een gesprek, impliceren de resultaten een 5,6 procentpunt hogere kans voor deze groep om niet uitgenodigd te worden en een 3,5 procentpunt lagere kans om wel uitgenodigd te worden.

We vinden ook een indicatie voor een negatieve bias ten opzichte van kandidaten met Turks als tweede moedertaal, maar zoals weergegeven in Tabel 3.4 is dit alleen het geval met betrekking tot de vraag of de kandidaat niet past bij de organisatie en is dit alleen significant bij een significantieniveau van tien procent. Al met al stellen we vast dat Marokkaanse kandidaten in het bijzonder, en meer specifiek mannelijke Marokkaanse kandidaten, minder positief beoordeeld worden.

Tabel 3.4 Vooral Marokkaanse mannen worden negatiever beoordeeld, maar de gestructureerde beoordelingsmethode verhelpt dit deels

	Beoordeling	Match met organisatie		Uitnodigen voor een gesprek	
		Past niet goed	Past goed	Nee	Ja
Tweede moedertaal (zie Tabel 3.3)					
Tweede moedertaal	-0,033 (0,034)	0,515** (0,204)	-0,218* (0,127)	0,506*** (0,192)	-0,241* (0,135)

Tweede moedertaal * treatment	-0,013 (0,048)	-0,365 (0,282)	-0,110 (0,180)	-0,457* (0,267)	-0,090 (0,190)
Specifieke tweede moedertaal					
Arabisch (Marokkaans)	-0,043 (0,042)	0,659*** (0,244)	-0,327** (0,157)	0,730*** (0,230)	-0,390** (0,167)
Arabisch (Marokkaans) * treatment	-0,008 (0,060)	-0,692** (0,341)	0,056 (0,220)	-0,842*** (0,323)	0,174 (0,234)
Turks	-0,022 (0,042)	0,423* (0,251)	-0,124 (0,154)	0,324 (0,238)	-0,034 (0,166)
Turks * treatment	-0,018 (0,059)	-0,269 (0,350)	-0,284 (0,220)	-0,067 (0,329)	-0,450* (0,233)
Specifieke tweede moedertaal, per geslacht					
Arabisch (Marokkaans) * man	-0,090 (0,061)	1,215*** (0,349)	-0,372* (0,223)	1,428*** (0,331)	-0,450** (0,237)
Arabisch (Marokkaans) * man * treatment	0,038 (0,085)	-1,413*** (0,489)	0,076 (0,313)	-1,822*** (0,468)	0,570* (0,335)
Arabisch (Marokkaans) * vrouw	0,003 (0,060)	0,306 (0,348)	-0,264 (0,223)	0,346 (0,327)	-0,158 (0,238)
Arabisch (Marokkaans) * vrouw * treatment	-0,053 (0,085)	-0,294 (0,484)	0,046 (0,314)	-0,336 (0,460)	-0,422 (0,332)
Turks * man	0,047 (0,060)	0,419 (0,368)	-0,019 (0,220)	0,309 (0,354)	-0,112 (0,240)
Turks * man * treatment	-0,087 (0,086)	-0,407 (0,520)	-0,463 (0,315)	-0,029 (0,491)	-0,375 (0,339)
Turks * vrouw	-0,098 (0,062)	0,518 (0,358)	-0,222 (0,225)	0,294 (0,342)	-0,008 (0,237)
Turks * vrouw * treatment	0,059 (0,085)	-0,369 (0,484)	-0,148 (0,316)	-0,014 (0,460)	-0,478 (0,327)

Bron: SEO Economisch Onderzoek (2023). n vignetten = 5.352, n respondenten = 446, * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$. Gecorrigeerd voor leeftijd, geslacht, opleidingsniveau, werkervaring, stagebeoordeling, Engelse taalvaardigheden en een dummy voor treatment (zie Tabel 3.3)

3.4.2 Effect van de gestructureerde beoordelingsmethode voor subgroepen

Ook bij deze modellen geeft de interactieterm tussen de tweede moedertaal en de treatment dummy het extra effect van de gestandaardiseerde en gestructureerde beoordelingsmethode weer voor deze specifieke groep. De analyses laten ook voor de specifieke groepen zien dat de gestandaardiseerde en gestructureerde beoordelingsmethode de bias gedeeltelijk kan verhelpen. Dat wil zeggen, zowel voor de kans dat een kandidaat niet als goede match met de organisatie wordt beschouwd als voor de kans dat een kandidaat niet wordt uitgenodigd voor een gesprek, vinden we significante effecten van de interactieterm tussen de treatment en de dummy voor Marokkaanse (mannelijke) kandidaten. De bias ten opzichte van Marokkaanse kandidaten - zoals blijkt uit de significante coëfficiënt voor de dummy voor Marokkaanse kandidaten - wordt gecompenseerd door het effect in omgekeerde richting van de interactie tussen de treatment en de dummy voor Marokkaanse kandidaten (zie Tabel 3.4). Hetzelfde geldt voor de subgroep van Marokkaanse mannelijke kandidaten. Zoals weergegeven in Tabel 3.2, met de berekende kansen voor de gemiddelde kandidaat, daalt de kans om niet uitgenodigd te worden voor een gesprek voor de gemiddelde Marokkaanse mannelijke kandidaat van 0,31 procent (conventionele methode)

naar 0,05 procent (gestructureerde methode), waar die voor de gemiddelde mannelijke kandidaat zonder tweede moedertaal gelijk blijft op 0,07 procent voor beide methodes. De kans om niet te worden gezien als goede match met de organisatie daalt voor de gemiddelde Marokkaanse mannelijke kandidaat van 0,25 procent (conventionele methode) naar 0,06 procent (gestructureerde methode), waar die voor de gemiddelde mannelijke kandidaat zonder tweede moedertaal gelijk blijft op 0,07 procent voor beide methodes.

Uit de resultaten blijkt bovendien een vergelijkbaar effect met betrekking tot de bereidheid om een kandidaat wel uit te nodigen voor een gesprek. Zoals te zien in Tabel 3.2 met de berekende kansen voor de gemiddelde kandidaat stijgt de kans om uitgenodigd te worden voor een gesprek voor de gemiddelde Marokkaanse mannelijke kandidaat van 93 procent (conventionele methode) naar 96 procent (gestructureerde methode), waar die voor de gemiddelde mannelijke kandidaat zonder tweede moedertaal gelijk blijft op 96 procent voor beide methodes.

Hoewel er dus voor de gehele onderzoeksgroep beperkte effecten bestaan van het wisselen van beoordelingsmethode, laten de resultaten voor specifieke subgroepen zien dat een negatieve bias (gedeeltelijk) kan worden gecompenseerd door het inzetten van een gestandaardiseerde en gestructureerde beoordelingsmethode.

In aanvulling op bovenstaande modellen hebben we ook getest of er, naast geslacht, andere groepsspecifieke effecten zijn, door interacties met de andere attributen mee te nemen. De resultaten van deze modellen wijzen niet op andere verschillen in de beoordeling tussen (sub)groepen van kandidaten.

3.5 Robuustheid van de resultaten

Er zijn verschillende analyses uitgevoerd om de robuustheid van de resultaten te toetsen. Door verschillende regressiemodellen te gebruiken en door de gebruikte modellen ook te schatten met alleen de gegevens van de eerste beoordelingsronde in het experiment krijgen we een goed beeld van het gedrag van recruiters gedurende het experiment en van de kwaliteit en reikwijdte van de data voor de uitgevoerde analyses. Deze robuustheidsanalyses bevestigen onze conclusies en leiden niet tot kwalitatief en kwantitatief andere bevindingen.

Tot slot is er in het experiment een aantal stellingen opgenomen om de neiging tot sociaalwenselijke antwoorden van recruiters te toetsen. Het zou namelijk zo kunnen zijn dat sommige recruiters op een sociaalwenselijke manier de vragen invullen en/of het experiment ondergaan, wat onze resultaten zou vertekenen. Bijlage B toont de resultaten wanneer we corrigeren voor mogelijke bias als gevolg van sociaalwenselijk gedrag. De geschatte coëfficiënten veranderen niet substantieel, wat betekent dat de uitkomsten van het experiment niet vertekend worden door sociaalwenselijk antwoordgedrag van de recruiters.

4 Impact

Er is op allerlei vlakken in de samenleving sprake van biases. Dit is het gevolg van impliciete of expliciete keuzes die leiden tot bepaald gedrag. Het verminderen van biases leidt tot meer kansengelijkheid. Ook onder recruiters lijkt sprake van biases bij het selecteren van kandidaten voor mbo-stages.

De bijdrage van dit onderzoek is een empirische analyse naar biases waaruit blijkt dat een alternatieve wijze van het presenteren en beoordelen van cv's tijdens de selectie van kandidaten voor een stage helpt bij het verminderen van een negatieve bias ten aanzien van kandidaten met een migratieachtergrond. De resultaten van het onderzoek laten zien dat kandidaten met een tweede moedertaal minder positief beoordeeld worden dan kandidaten zonder tweede moedertaal. Dit geldt vooral voor de groep mannelijke Marokkaanse kandidaten. De empirische analyses wijzen erop dat de kans dat de gemiddelde Marokkaanse mannelijke kandidaat wordt gezien als goede match met de organisatie acht procentpunt lager is dan voor de gemiddelde mannelijke kandidaat zonder tweede moedertaal. De kans dat de gemiddelde Marokkaanse mannelijke kandidaat wordt uitgenodigd voor een gesprek is drie procentpunt lager dan voor de gemiddelde mannelijke kandidaat zonder tweede moedertaal.

De resultaten van deze studie laten zien dat de verschillen tussen stagekandidaten met een tweede moedertaal en kandidaten zonder tweede moedertaal deels verdwijnen door een meer gestructureerde beoordelingsmethode toe te passen. Dit effect is niet in alle gevallen statistisch significant, maar bij de gestructureerde methode is de kans dat de gemiddelde Marokkaanse mannelijke kandidaat wordt uitgenodigd voor een gesprek even groot als de kans dat de gemiddelde mannelijke kandidaat zonder tweede moedertaal wordt uitgenodigd. Hiermee wordt een groot deel van de bias voor deze groep tenietgedaan, wat hun kansen bevordert.

De resultaten zijn beleidsrelevant omdat er invulling wordt gegeven aan het gevoel dat studenten met een migratieachtergrond gemiddeld genomen meer moeite hebben met het vinden van een stageplaats. Uit eerder onderzoek komt naar voren dat mbo-studenten met een niet-westerse migratieachtergrond vaker moeten solliciteren om een stage te vinden dan mbo-studenten zonder migratieachtergrond (ROA, 2018). Dit onderzoek laat zien dat er ook bij correctie voor andere factoren, zoals werkervaring, opleidingsniveau of leeftijd, sprake is van biases bij de selectie van stagekandidaten. De impact van die uitkomsten is dat bij de conventionele cv-beoordeling bepaalde mbo-studenten een minder grote kans hebben op een stage van hun voorkeur. Het alternatief zal vaak minder aantrekkelijk zijn, waardoor de kwaliteit van de stage lager is. Een stage van een lagere kwaliteit heeft een negatieve impact op de kans op een goede start op de arbeidsmarkt (Bisschop et al., 2020 en Van den Berg et al., 2020). Daarnaast heeft de negatieve ervaring met het vinden van een stage ook negatieve effecten op het welbevinden van studenten (Andriessen et al., 2021) en kan het demotiverend werken voor toekomstige keuzes op de arbeidsmarkt. Om gelijkere kansen te realiseren is het daarom van belang om biases bij het selecteren van kandidaten voor mbo-stages zoveel mogelijk te verminderen.

De resultaten van het onderzoek zijn ook relevant voor recruitment. Dit onderzoek laat zien dat wanneer eerst de ervaring en competenties van kandidaten worden beoordeeld en daarna pas het volledige cv, de bias in de selectie van kandidaten kan dalen. Deze methode is relatief eenvoudig te implementeren en doet niets af aan de totale hoeveelheid informatie die werkgevers ontvangen over de kandidaten. Omdat de meer gestructureerde methode naar verwachting eenvoudig is toe te passen en geen grote investeringen of aanpassingen van het selectieproces

vereist, is deze methode ook bruikbaar voor organisaties van een kleinere omvang of zonder uitgebreide HR-afdeling.

Een kanttekening bij de brede toepassing van deze methode is dat het onderzoek bestaat uit een experiment met fictieve sollicitanten voor een fictieve vacature. Voordat deze gestandaardiseerde en gestructureerde beoordelingsmethode in de praktijk wordt toegepast is het noodzakelijk om nader onderzoek uit te voeren naar deze veelbelovende methode. Een dergelijk onderzoek zou zich moeten richten op de externe validiteit van deze methode voor het selecteren van stagiairs in het mbo. Externe validiteit gaat over de mate waarin de bevindingen en conclusies van dit onderzoek kunnen worden gegeneraliseerd of toegepast op de totale populatie van studenten en de mate waarin ze geldig zijn in alle omgevingen en omstandigheden buiten deze specifieke context. Met andere woorden, het is de mate waarin de resultaten verkregen uit deze steekproef van recruiters representatief kunnen worden geacht voor het gehele mbo en toepasbaar zijn in praktische situaties. In gesprekken met het mbo en leerbedrijven zou dat verder kunnen worden uitgediept om tot een beoordelingsmethode te komen die haalbaar en effectief is in de praktijk.

Ten slotte vergroot dit onderzoek het bewustzijn over negatieve biases in het mbo en draagt het bij aan het stimuleren van het maatschappelijke debat over kansengelijkheid. De resultaten van het onderzoek bieden nieuwe perspectieven op sociale ongelijkheid, wat kan leiden tot veranderingen in houdingen, gedrag of beleidsbeslissingen. We vinden dat er in ons onderzoek sprake is van negatieve biases tegenover kandidaten met een migratieachtergrond, maar ook dat dit vermindert als er meer gefocust wordt op de competenties van een kandidaat. Dit inzicht kan gebruikt worden om het sollicitatieproces te verbeteren en zo gelijke kansen voor mbo-studenten te stimuleren, maar het kan ook gebruikt worden om bewustzijn te creëren bij bijvoorbeeld werkgevers. Bewuster stilstaan bij de relevante vaardigheden van een kandidaat kan het effect van een (al dan niet onbewust aanwezige) bias al sterk verminderen voor de groepen jongeren die het meest last hebben van dergelijke biases.

Referenties

- Aigner, D. J., & Cain, G. G. (1977). Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review*, 30(2), 175-187.
- Andriessen, I., Van Rooijen, M., Day, M., Van den Berg, A., Mienis, E. & Verweij, N. (2021). *Ongelijke kansen op de stagemarkt. Onderzoek naar objectief vastgestelde en ervaren stagediscriminatie in het mbo in Utrecht*. Verwey-Jonker Instituut.
- Arrow, K. J. (1998). What Has Economics to Say About Racial Discrimination? *Journal of Economic Perspectives*, 12(2), 91-100.
- Becker, G. (1971). *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Van den Berg, E., Bisschop, P., Ter Weel, B. & Zwetsloot, J. (2020). *Verschillen in arbeidsparticipatie van mbo'ers circa een jaar na afstuderen. Analyse van de verschillen tussen jongeren met een verschillende achtergrond op basis van administratieve data en enquêtedata*. SEO Economisch Onderzoek.
- Bisschop, P., Ter Weel, B., & Zwetsloot, J. (2020). Ethnic employment gaps of graduates in the Netherlands. *De Economist*, 168(4), 577-598.
- Blommaert, L., Van Tubergen, F., Coenders, M. (2012). Implicit and explicit interethnic attitudes and ethnic discrimination in hiring. *Soc Sci Res.*, 41(1), 61-73.
- Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and prejudice: The role of value conflicts. *Psychological science*, 24(2), 140-149.
- Derous, E., & Ryan, A. M. (2019). When your resume is (not) turning you down: Modelling ethnic bias in resume screening. *Human Resource Management Journal*, 29(2), 113-130.
- Dagevos, J., De Voogd-Hamelink, M., & Damen, R. (2022). *Gevestigd, maar niet thuis*. Den Haag: SCP.
- Dolfing, M., & Van Tubergen, F.. (2005). Bensaïdi of Veenstra? Een experimenteel onderzoek naar discriminatie van Marokkanen in Nederland. *Sociologie*, 1(4), 407-422.
- Dovidio, J. F., & Gaertner, S. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, 11(4), 315-319.
- Hagendoorn, L. (1995). Intergroup Biases in Multiple Group Systems: The Perception of Ethnic Hierarchies. *European Review of Social Psychology*, 6(1), 199-228.
- Kaas, L., & Manger, C. (2012). Ethnic discrimination in Germany's labour market: A field experiment. *German Economic Review*, 13(1), 1-20.
- Lancee, B. (2021). Ethnic discrimination in hiring: comparing groups across contexts. Results from a cross-national field experiment. *Journal of Ethnic and Migration Studies*, 47(6), 1181-1200.

- Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research*, 61(5), 534-547.
- Lippens, L., Vermeiren, S., & Baert, S. (2023). The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments. *European Economic Review*, 151, 104315.
- Navarrette, C. D., McDonald, M. M., & Sidanius, J. (2010). Prejudice at the Nexus of Race and Gender: An Outgroup Male Target Hypothesis. *Journal of Personality and Social Psychology*, 98(6), 933-945.
- Nosek, B.A., 2007. Implicit-explicit relations. *Current Directions in Psychological Science*, 16(2), 65-69.
- Pettigrew, T. F., & Tropp, L. R. (2011). *When groups meet: The dynamics of intergroup contact*. New York: Psychology Press.
- Phelps, E. S. (1972). The Statistical theory of Racism and Sexism. *American Economic Review*, 62(4), 659-661.
- ROA (2018). *MBO Stages en de migratieachtergrond van studenten* (ROA-F-2018-/17). Researchcentrum voor Onderwijs en Arbeidsmarkt.
- Rooth, D-O. (2010). Automatic Associations and Discrimination in Hiring: Real World Evidence. *Labour Economics*, 17(3), 523-534.
- Scheepers, P., Gijsberts, M., & Coenders, M. (2002). Ethnic Exclusionism in European Countries : Public Opposition to Civil Rights for Legal Migrants as a Response to Perceived Ethnic Threat. *European Sociological Review*, 18(1), 17-34.
- Sidanius, J., Pratto, F., Van Laar, C., & Levin, S. (2004). Social dominance theory: Its agenda and method. *Political Psychology*, 25(6), 845-880.
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28(2), 191-193.
- Thijssen, L., Van Tubergen, L., Coenders, M., Hellpap, R. , & Jak, S. (2022). Discrimination of Black and Muslim Minority Groups in Western Societies: Evidence From a Meta-Analysis of Field Experiments. *International Migration Review*, 56(3), 843-880.
- 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.
- Vernby, K. , & Dancygier, R. (2019). Can immigrants counteract employer discrimination? A factorial field experiment reveals the immutability of ethnic hierarchies. *Plos one*, 14(7), e0218044.
- Wolgast, S., Backstrom, M. , & Bjorklund, F. (2017). Tools for fairness: Increased structure in the selection process reduces discrimination. *Plos one*, 12(12).

Bijlage A Linear random intercept en OLS modellen

Tabel A.1 toont de geschatte coëfficiënten van een linear random intercept model. Dit model maakt een directere interpretatie van de coëfficiënten mogelijk en wordt besproken in paragraaf 3.3.1. Het model is vergelijkbaar met het random intercept model in Tabel 3.3. Tabel A.2 schat hetzelfde model voor de specificatie in Tabel 3.4 in de hoofdttekst. Tabel A.3 presenteert ten slotte de schattingen van een OLS-model (specificatie zoals in Tabel 3.4 in de hoofdttekst).

Tabel A.1 Regressieresultaten van het linear random intercept model

	Match met organisatie		Uitnodigen voor een gesprek	
	Past niet goed	Past goed	Nee	Ja
Talenkennis (naast Engels) (geen tweede moedertaal als referentiecategorie)				
Tweede moedertaal (Turks of Arabisch (Marokkaans))	0,020** (0,009)	-0,018 (0,013)	0,024** (0,009)	-0,015 (0,013)
Tweede moedertaal * Treatment	-0,014 (0,013)	-0,018 (0,019)	-0,017 (0,013)	-0,018 (0,018)
Leeftijd (gemiddeld als referentiecategorie)				
Relatief oud	-0,003 (0,006)	0,002 (0,009)	-0,000 (0,006)	-0,000 (0,009)
Geslacht (man als referentiecategorie)				
Vrouw	-0,003 (0,006)	0,005 (0,009)	-0,005 (0,006)	0,009 (0,009)
Opleidingsniveau (niveau 2 als referentiecategorie)				
Mbo-niveau 3	-0,087*** (0,009)	0,133*** (0,013)	-0,092*** (0,009)	0,131*** (0,013)
Mbo-niveau 4	-0,107*** (0,009)	0,186*** (0,013)	-0,112*** (0,009)	0,171*** (0,012)
Werkervaring (geen relevante werkervaring als referentiecategorie)				
1 jaar	-0,032*** (0,008)	0,108*** (0,011)	-0,030*** (0,008)	0,093*** (0,011)
2 jaar	-0,041*** (0,008)	0,145*** (0,011)	-0,041*** (0,008)	0,113*** (0,011)
Stagebeoordeling (geen beoordeling vermeld als referentiecategorie)				
1 positieve beoordeling	-0,044*** (0,010)	0,114*** (0,015)	-0,057*** (0,010)	0,115*** (0,014)
2 positieve beoordelingen	-0,063*** (0,011)	0,195*** (0,016)	-0,069*** (0,011)	0,158*** (0,015)
1 zeer positieve beoordeling	-0,064*** (0,010)	0,174*** (0,015)	-0,074*** (0,010)	0,150*** (0,014)
2 zeer positieve beoordelingen	-0,056*** (0,011)	0,195*** (0,016)	-0,067*** (0,011)	0,159*** (0,015)
Talenkennis Engels (gemiddeld als referentiecategorie)				

Goed	0,003 (0,006)	0,009 (0,009)	-0,001 (0,007)	0,002 (0,009)
Treatment	0,012 (0,009)	0,011 (0,013)	0,014 (0,009)	-0,006 (0,012)
Constante	0,221*** (0,015)	0,248*** (0,024)	0,242*** (0,015)	0,472*** (0,021)

Bron: SEO Economisch Onderzoek (2023). *n* vignetten = 5,352, *n* respondenten = 446, **p*<0,10, ***p*<0,05, ****p*<0,01

Tabel A.2 Regressieresultaten van het linear random intercept model voor specifieke subgroepen

	Match met organisatie		Uitnodigen voor een gesprek	
	Past niet goed	Past goed	Nee	Ja
Tweede moedertaal				
Tweede moedertaal	0,020** (0,009)	-0,018 (0,013)	0,024** (0,009)	-0,015 (0,013)
Tweede moedertaal * treatment	-0,014 (0,013)	-0,018 (0,019)	-0,017 (0,0130)	-0,018 (0,018)
Specifieke tweede moedertaal				
Arabisch (Marokkaans)	0,029*** (0,011)	-0,028* (0,016)	0,034*** (0,012)	-0,029* (0,015)
Arabisch (Marokkaans) * treatment	-0,026*** (0,016)	-0,005 (0,023)	-0,035** (0,016)	0,005 (0,022)
Turks	0,011 (0,011)	-0,008 (0,016)	0,014 (0,012)	-0,021 (0,015)
Turks * treatment	-0,001 (0,016)	-0,031 (0,023)	0,004 (0,016)	-0,041* (0,022)
Specifieke tweede moedertaal, per geslacht				
Arabisch (Marokkaans) * man	0,040** (0,016)	-0,036 (0,024)	0,056*** (0,017)	-0,035 (0,022)
Arabisch (Marokkaans) * man * treatment	-0,043* (0,022)	0,000 (0,033)	-0,070*** (0,023)	0,040 (0,022)
Arabisch (Marokkaans) * vrouw	0,018 (0,016)	-0,019 (0,023)	0,012 (0,016)	-0,023 (0,022)
Arabisch (Marokkaans) * vrouw * treatment	-0,009 (0,022)	-0,010 (0,033)	0,000 (0,023)	-0,031 (0,031)
Turks * man	-0,008 (0,016)	0,010 (0,010)	0,007 (0,016)	0,005 (0,022)
Turks * man * treatment	0,015 (0,022)	-0,054 (0,033)	0,001 (0,023)	-0,041 (0,031)
Turks * vrouw	0,032* (0,016)	-0,028 (0,024)	0,023 (0,017)	-0,009 (0,023)
Turks * vrouw * treatment	-0,020 (0,022)	-0,007 (0,033)	-0,003 (0,023)	-0,040 (0,031)

Bron: SEO Economisch Onderzoek (2023). *n* vignetten = 5.352, *n* respondenten = 446, **p*<0,10, ***p*<0,05, ****p*<0,01. Gecorrigeerd voor leeftijd, geslacht, opleidingsniveau, werkervaring, stagebeoordeling, Engelse taalvaardigheden en een dummy voor treatment (zie Tabel 3.3)

Tabel A.3 Regressieresultaten van het OLS-model voor specifieke subgroepen

	Beoordeling	Match met organisatie		Uitnodigen voor een gesprek	
		Past niet goed	Past goed	Nee	Ja
Tweede moedertaal					
Tweede moedertaal	-0,065 (0,050)	0,236* (0,135)	-0,058 (0,082)	0,274** (0,130)	-0,095 (0,094)
Tweede moedertaal * treatment	0,035 (0,070)	-0,219 (0,189)	-0,056 (0,115)	-0,255 (0,181)	-0,063 (0,131)
Specifieke tweede moedertaal					
Arabisch (Marokkaans)	-0,109* (0,061)	0,343** (0,162)	-0,103 (0,101)	0,390** (0,155)	-0,177 (0,114)
Arabisch (Marokkaans) * treatment	0,069 (0,086)	-0,288 (0,228)	0,016 (0,142)	-0,361* (0,219)	0,054 (0,161)
Turks	-0,021 (0,061)	0,126 (0,167)	-0,013 (0,100)	0,154 (0,160)	-0,013 (0,115)
Turks * treatment	0,002 (0,086)	-0,148 (0,234)	-0,127 (0,142)	-0,145 (0,223)	-0,179 (0,161)
Specifieke tweede moedertaal, per geslacht					
Arabisch (Marokkaans) * man	-0,292*** (0,087)	0,575** (0,224)	-0,288** (0,143)	0,670*** (0,217)	-0,371** (0,162)
Arabisch (Marokkaans) * man * treatment	0,253** (0,122)	-0,620* (0,320)	0,269 (0,200)	-0,837*** (0,308)	0,467** (0,227)
Arabisch (Marokkaans) * vrouw	0,070 (0,086)	0,104 (0,236)	0,080 (0,142)	0,077 (0,224)	0,013 (0,162)
Arabisch (Marokkaans) * vrouw * treatment	-0,110 (0,123)	0,053 (0,328)	-0,240 (0,202)	0,133 (0,316)	-0,372 (0,229)
Turks * man	-0,027 (0,085)	0,111 (0,243)	0,021 (0,141)	0,266 (0,231)	-0,082 (0,163)
Turks * man * treatment	0,025 (0,123)	-0,181 (0,345)	-0,134 (0,202)	-0,411 (0,328)	0,018 (0,231)
Turks * vrouw	-0,022 (0,087)	0,139 (0,232)	-0,055 (0,143)	0,048 (0,224)	0,054 (0,163)
Turks * vrouw * treatment	-0,014 (0,122)	-0,126 (0,321)	-0,113 (0,199)	0,104 (0,309)	-0,377* (0,226)

Bron: SEO Economisch Onderzoek (2023). n vignetten = 5.352, n respondenten = 446, * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$. Gecorrigeerd voor leeftijd, geslacht, opleidingsniveau, werkervaring, stagebeoordeling, Engelse taalvaardigheden en een dummy voor treatment (zie Tabel 3.3).

Bijlage B Controle voor sociaalwenselijkheid

Mogelijk geven sommige recruiters tijdens het experiment sociaalwenselijke antwoorden. Om te proberen hiervoor te controleren, zijn alle recruiters tien stellingen voorgelegd over de sociaalwenselijkheid van hun gedrag. De stellingen zijn afkomstig uit de persoonlijkheidsvragenlijst van de LISS Core Study en zijn een Nederlandse vertaling van een verkorte vorm van de Marlowe-Crowne social desirability scale, opgesteld door Strahan en Gerbasi (1972). Mogelijke antwoorden varieerden van 1 (dit is helemaal niet op mij van toepassing) tot 5 (dit is helemaal op mij van toepassing), waarbij 4 en 5 werden beschouwd als een sociaalwenselijk antwoord. Om te controleren voor bias door sociaalwenselijke antwoorden, hebben we een dummy opgenomen in de regressies, die gelijk is aan 1 als het aantal sociaalwenselijke antwoorden boven het gemiddelde ligt en 0 als dat niet zo is.

Tabel B.1 geeft de geschatte coëfficiënten van het random intercept model zoals in tabel 3 wanneer we controleren voor onze maatstaf voor sociaalwenselijkheid. De resultaten laten geen significante verschillen zien tussen de geschatte coëfficiënten in Tabel 3.4 in de hoofdtekst en Tabel B.1.

Andere maatstaven voor sociaalwenselijkheid, waarbij een grotere afwijking van het gemiddelde zwaarder weegt, zoals het kwadraat van het aantal sociaalwenselijke antwoorden of de log functie (zoals voorgesteld door Larson, 2019), leiden tot vergelijkbare conclusies.

Tabel B.1 Regressieresultaten bij correctie voor mogelijke sociaalwenselijke antwoorden

	Beoordeling	Match met organisatie		Uitnodigen voor een gesprek	
		Past niet goed	Past goed	Nee	Ja
Tweede moedertaal					
Tweede moedertaal	-0,040 (0,044)	0,888*** (0,272)	-0,253 (0,162)	0,770*** (0,248)	-0,354** (0,177)
Tweede moedertaal * treatment	-0,013 (0,048)	-0,612** (0,281)	-0,155 (0,175)	-0,562** (0,266)	-0,106 (0,190)
Specifieke tweede moedertaal					
Arabisch (Marokkaans)	-0,016 (0,055)	0,910*** (0,328)	-0,368* (0,205)	0,776** (0,308)	-0,409* (0,220)
Arabisch (Marokkaans) * treatment	-0,008 (0,060)	-0,719** (0,336)	0,072 (0,221)	-0,963*** (0,324)	0,195 (0,234)
Turks	-0,063 (0,055)	0,930*** (0,329)	-0,302 (0,204)	0,903*** (0,303)	-0,289 (0,217)
Turks * treatment	-0,014 (0,059)	-0,185 (0,345)	-0,320 (0,221)	-0,100 (0,329)	-0,351 (0,233)
Specifieke tweede moedertaal, per geslacht					
Arabisch (Marokkaans) * man	-0,021 (0,078)	0,665 (0,486)	-0,496* (0,291)	1,083** (0,458)	-0,414 (0,317)
Arabisch (Marokkaans) * man * treatment	0,040 (0,085)	-0,865* (0,481)	0,147 (0,315)	-1,762*** (0,475)	0,612* (0,336)
Arabisch (Marokkaans) * vrouw	-0,014 (0,077)	0,762 (0,453)	-0,165* (0,289)	0,201 (0,413)	-0,318 (0,310)
Arabisch (Marokkaans) * vrouw * treatment	-0,054	-0,323	-0,006	-0,206	-0,319

	(0,085)	(0,486)	(0,314)	(0,465)	(0,333)
Turks * man	0,043 (0,079)	0,229 (0,510)	-0,275 (0,295)	0,581 (0,472)	-0,188 (0,324)
Turks * man * treatment	-0,085 (0,086)	0,519 (0,514)	-0,439 (0,317)	-0,011 (0,493)	-0,363 (0,340)
Turks * vrouw	-0,166** (0,078)	0,726 (0,462)	-0,233 (0,291)	1,076*** (0,413)	-0,322 (0,302)
Turks * vrouw * treatment	0,050 (0,085)	-0,202 (0,490)	-0,197 (0,318)	-0,188 (0,460)	-0,433 (0,329)

Bron: SEO Economisch Onderzoek (2023). n vignetten = 5.352, n respondenten = 446, * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$. Gecorrigeerd voor leeftijd, geslacht, opleidingsniveau, werkervaring, stagebeoordeling, Engelse taalvaardigheden en een dummy voor treatment (zie Tabel 3.3) en een dummy voor sociaalwenselijkheid



“De wetenschap dat het goed is.”

SEO Economisch Onderzoek doet onafhankelijk toegepast onderzoek in opdracht van overheid en bedrijfsleven. Ons onderzoek helpt onze opdrachtgevers bij het nemen van beslissingen. SEO Economisch Onderzoek is gelieerd aan de Universiteit van Amsterdam. Dat geeft ons zicht op de nieuwste wetenschappelijke methoden. We hebben geen winstoogmerk en investeren continu in het intellectueel kapitaal van de medewerkers via promotietrajecten, het uitbrengen van wetenschappelijke publicaties, kennisnetwerken en congresbezoek.

SEO-rapport 2023-90
ISBN 978-90-5220-318-8

Informatie & Disclaimer

SEO Economisch Onderzoek heeft op de verkregen informatie en data geen onderzoek uitgevoerd dat het karakter draagt van een accountantscontrole of due diligence. SEO is niet verantwoordelijk voor fouten of omissies in de verkregen informatie en data.

Copyright © 2023 SEO Amsterdam.

Alle rechten voorbehouden. Het is geoorloofd gegevens uit dit rapport te gebruiken in artikelen, onderzoeken en collegesyllabi, mits daarbij de bron duidelijk en nauwkeurig wordt vermeld. Gegevens uit dit rapport mogen niet voor commerciële doeleinden gebruikt worden zonder voorafgaande toestemming van de auteur(s). Toestemming kan worden verkregen via secretariaat@seo.nl.

Roetersstraat 29
1018 WB Amsterdam

+31 20 399 1255
secretariaat@seo.nl
www.seo.nl