

Praktijktesten: van onderzoeksmethode naar beleidsinstrument tegen discriminatie?

*Pieter-Paul Verhaeghe^{*1,2} & Koen Van der Bracht²*

Samenvatting

De voorbije jaren kwamen praktijktesten uitgebreid in de media. Academici brachten grootschalige discriminatie op de woning- en arbeidsmarkt aan het licht door middel van praktijktesten. Verschillende middenveldorganisaties en politici pleiten dan weer om praktijktesten te gebruiken om discriminerende bedrijven of verhuurders aan te pakken. Dit overzichtsartikel wil een brug slaan tussen beide werelden en nagaan in welke mate praktijktesten gebruikt kunnen worden als wetenschappelijke methode én als beleidsinstrument in de strijd tegen discriminatie. We maken eerst op basis van drie criteria een typologie op van de technieken om discriminerend gedrag te testen: situatietesten, correspondentietesten, mystery calls/visits en mystery mails. We staan hierbij telkens stil bij hun ontstaansgeschiedenis, toepassing en voor- en nadelen. Deze typologie is zeker niet algemeen aanvaard, maar ze kan toekomstig onderzoek conceptueel sturen. Vervolgens bespreken we hoever het onderzoek naar praktijktesten als beleidsinstrument staat. Er is beperkte wetenschappelijke evidentie dat praktijktesten in bepaalde omstandigheden en onder bepaalde voorwaarden effectief, haalbaar en ethisch verantwoord zijn. Wij zijn daarom voorzichtige voorstanders van het gebruik van praktijktesten als beleidsinstrument tegen discriminatie.

Kernwoorden

praktijktesten, mystery shopping, discriminatie, woningmarkt, arbeidsmarkt, beleid, methodologie

* pieter-paul.verhaeghe@vub.ac.be

1 Vakgroep Sociologie, Vrije Universiteit Brussel

2 Vakgroep Sociologie, Universiteit Gent

Inleiding

Discriminatie verwijst naar de nadelige behandeling van mensen omwille van hun etnisch-culturele afkomst, gender, leeftijd, seksuele geaardheid, handicap of een ander beschermd kenmerk. Reeds verschillende overzichtswerken hebben structurele discriminatie aangetoond op de arbeids- en woningmarkt (Pager & Shepherd, 2008; Riach & Rich, 2002; Rich, 2014; Zschirnt & Ruedin, 2016). Ook in België stellen we vast dat etnische minderheden (Capéau, Eeman, Groenez & Lamberts, 2012; Heylen & Van den Broeck, 2015; Van der Bracht, Coenen & Van de Putte, 2015; Van der Bracht & Verhaeghe, 2016a), vrouwen (Baert, De Pauw & Deschacht, 2016), ouderen (Capéau et al., 2012) en personen met een handicap (Verhaeghe, Van der Bracht & Van de Putte, 2016) nadeliger behandeld worden. De discriminatie gebeurt zowel door institutionele spelers (bv. makelaars of interimkantoren) als door particulieren.

Discriminatie is ingebed in een bredere samenleving die het zowel kan beteugelen als aanwakkeren. We stellen vast dat zowel de attitudes over bijvoorbeeld migranten (Ceobanu & Escandell, 2010; Coenders, Lubbers, Scheepers & Verkuyten, 2008; Ford, 2008; Meuleman, Davidov & Billiet, 2009; Semyonov & Gorodzeisky, 2006) als het discriminerend gedrag (Turner, Santos, Levy, Wissoker, Aranda & Pitingolo, 2013; Zschirnt & Ruedin, 2016) sterk verschillen tussen sociale groepen, tussen landen en doorheen de tijd. Er bestaat heel wat debat over wat precies het niveau en de aard van discriminatie bepaalt. Verschillende auteurs stellen dat het de bredere cultuur is die discriminatie aanwakkert, zoals ‘alledaags racisme’ (Essed, 1991), ‘institutioneel racisme’ (Özdil, 2015), ‘ocularcentrisme’ (Bolt, 2005) of ‘seksisme’ (Risman, 2004). Anderen verwijzen naar de invloed van de economische omstandigheden, zoals (stijgende) werkloosheid (Coenders & Scheepers, 2008; Meuleman et al., 2009) of de sociaaleconomische achterstelling van de buurt (Carlsson & Eriksson, 2015; Turner & Ross, 2005). Er zijn ten slotte ook auteurs die de impact van financiële bestraffing van discriminerende organisaties aantonen: na de boete gaat men minder discrimineren (Ross & Galster, 2007).

Het is tegen deze achtergrond dat men ook het debat over praktijktesten moet plaatsen. Verschillende middenveldorganisaties, politici, ambtenaren en onderzoekers bevelen het gebruik van praktijktesten als beleidsmiddel aan. De claim is dat de invoering van praktijktesten als beleidsinstrument een sterk wapen zou zijn in de strijd tegen discriminatie. Praktijktesten als onderzoeksmethode en praktijktesten als beleidsinstrument zijn echter twee verschillende zaken. Ze hebben andere doelstellingen en vergen ook een andere methodologische aanpak. Terwijl velen pleiten voor praktijktesten in het beleid, bestaat er nauwelijks onderzoek naar de effectiviteit, haalbaarheid en ethiek van praktijktesten als beleidsinstrument.

Dit overzichtsartikel wil een brug slaan tussen de onderzoeks- en beleidswereld en nagaan in welke mate praktijktesten gebruikt kunnen worden als beleidsinstrument in de strijd tegen discriminatie. We maken eerst een onderscheid tussen praktijktesten en mystery shopping en stellen vervolgens een typologie op van de technieken om discriminerend gedrag te testen. We staan hierbij telkens stil bij hun ontstaansgeschiedenis,

toepassing en voor- en nadelen. Vervolgens bespreken we hoever het onderzoek naar praktijktesten als beleidsinstrument staat. We eindigen met een reflectie over de wenselijkheid en haalbaarheid van praktijktesten als beleidsinstrument tegen discriminatie.

Het onderscheid tussen praktijktesten en mystery shopping

In het maatschappelijke debat bestaat er heel wat polemiek over het gebruik van praktijktesten en mystery shopping om discriminatie op te sporen. Het debat wordt helaas niet altijd even helder gevoerd, omdat verschillende concepten en methodes door elkaar gebruikt worden. Verschillende termen zoals praktijktesten, mystery shopping, gedragstesten en correspondentietesten passeren de revue. Er wordt ook nauwelijks een onderscheid gemaakt tussen de gebruikte methodes: het uitsturen van twee quasi-identieke e-mails met de vraag een huurwoning te mogen bezichtigen, kan niet vergeleken worden met het spelen van een discriminerende klant bij een dienstenchequebedrijf of met het betere acteerwerk door twee fictieve sollicitanten tijdens een sollicitatiegesprek. Het gebruik van praktijktesten en mystery shopping voor zowel wetenschappelijke als beleidsdoeleinden is gebaat met meer conceptuele klaarheid. Een duidelijke conceptualisatie van de verschillende testmethodes dringt zich dan ook op. We maken daarom in dit overzichtsartikel een conceptueel onderscheid tussen praktijktesten en mystery shopping en stellen vervolgens een typologie van discriminatietesten op. Dit onderscheid en deze typologie zijn niet algemeen aanvaard, maar ze hebben het voordeel van de duidelijkheid en kunnen het wetenschappelijke en maatschappelijke debat in de toekomst helpen stroomlijnen.

Praktijktesten

Praktijktesten zijn een experimentele methode, waarbij men nagaat in welke mate twee of meerdere quasi-identieke kandidaten ongelijk behandeld worden (Bertrand & Mullainathan, 2004; Fix & Turner, 1998; Pager & Western, 2012; Riach & Rich, 2002; Turner et al., 2013). Beide kandidaten solliciteren bijvoorbeeld voor een job, gaan in op een huuradvertentie of willen een discotheek binnengaan. De kandidaten verschillen niet van elkaar op zo veel mogelijk relevante kenmerken, behalve op het vlak van het te testen kenmerk dat zou kunnen aanleiding geven tot discriminatie (bv. huidskleur, gender, seksuele geaardheid...). Nadien gaat men na of een van de kandidaten nadeliger behandeld is geweest in vergelijking met de andere door bijvoorbeeld een werkgever, makelaar of portier. Men gaat er hierbij van uit dat een eventuele ongelijke behandeling tussen beide kandidaten te wijten is aan de discriminatiegrond. Cruciaal is dan ook dat de twee kandidaten zo gelijkaardig mogelijk zijn en enkel verschillen op het vlak van een specifiek beschermd kenmerk, zodat het onderscheid enkel en alleen toegeschreven kan worden aan discriminatie op basis van deze specifieke discriminatiegrond.

Het gebruik van praktijktesten gaat terug tot twee studies van Britse sociologen op het einde van de jaren zestig (Daniel, 1968; Jowell & Prescott-Clarke, 1970). Zij gebruikten paren van blanke en niet-blanke kandidaten om discriminatie op de woning- en arbeidsmarkt aan te tonen. Vanaf de jaren zestig werd de methode in toenemende mate gebruikt door onderzoekers, beleidsmakers en belangenorganisaties (Riach & Rich, 2002). Door middel van praktijktesten werd systematische discriminatie aangetoond in zowel de Verenigde Staten, Canada, Australië, China, India en verschillende Europese landen (voor twee excellente overzichtsstudies, zie Riach & Rich, 2002 en Rich, 2014).

In Europa en de Verenigde Staten bestaan er evenwel twee aparte tradities met betrekking tot het gebruik van praktijktesten (Zschirnt, 2016a). Een eerste verschilpunt is dat de praktijktesten in Europa voornamelijk gebeuren voor onderzoeksdoel-einden: het analyseren en verklaren van de oorzaken, patronen en gevolgen van discriminatie. In de Verenigde Staten daarentegen worden er praktijktesten uitgevoerd voor zowel onderzoeks- als juridische doeleinden. Woningagentschappen sturen bijvoorbeeld een blanke kandidaat-huurder naar een woning na een concrete klacht over discriminatie door een zwarte kandidaat-huurder (Yinger, 1998). Een speciale rol is daar weggelegd voor het door de overheid gesubsidieerde onderzoeksinstituut 'Urban Institute' dat lokale woningagentschappen ondersteunt in hun juridische praktijktesten en daarnaast circa tienjaarlijks het voorkomen van discriminatie op de woningmarkt monitort door middel van onderzoeksmatige praktijktesten. Een tweede verschil tussen beide continenten is dat de meeste praktijktesten in Europa discriminatie op de arbeidsmarkt onderzoeken, terwijl de Amerikaanse praktijktesten zich voornamelijk op de woningmarkt focussen. Zschirnt (2016a) verklaart dit vanuit de aansturing van het beleid. In het Verenigd Koninkrijk focuste de overheid zich hoofdzakelijk op de uitsluiting op de arbeidsmarkt (zie bv. Race Relations Acts van 1965 en 1976), terwijl in de Verenigde Staten voornamelijk de woningdiscriminatie een politiek heet hangijzer was (zie bv. de Fair Housing Act van 1968 ter uitbreiding van de Civil Rights Act).

In België dateert het gebruik van praktijktesten van de jaren negentig. Bovenkerk (1992) voerde in opdracht van de Internationale Arbeidsorganisatie een hele reeks van praktijktesten uit op de arbeidsmarkt. De meest recente praktijktesten worden in België uitgevoerd op de arbeidsmarkt (zoals Baert & Omeij, 2015; Baert, Cockx, Gheyle & Vandamme, 2015; Baert & Verhofstadt, 2015; Capéau et al., 2012) en de huisvestingsmarkt (zoals Heylen & Van den Broeck, 2015; Van der Bracht et al., 2015; Van der Bracht & Verhaeghe, 2016; Verhaeghe et al., 2016) voor louter onderzoeksdoeleinden, al dan niet in opdracht van een lokale, regionale of federale overheid. Recent gaan er stemmen op om ook in België praktijktesten als beleidsinstrument in te voeren en er zijn reeds een aantal lokale initiatieven, zoals praktijktesten op de huurwoningmarkt in opdracht van Stad Gent. Deze praktijktesten hebben een remediërend en juridisch doel, wat meer in de Amerikaanse dan Europese traditie past.

Mystery shopping

Mystery shopping is een onderzoeksmethode waarbij men aan een dienstverlenende organisatie een discriminerende vraag stelt. Men speelt bijvoorbeeld een verhuurder en vraagt aan een makelaar om de woning niet aan een persoon van vreemde origine te verhuren of men doet zich voor als een bedrijf en vraagt aan een uitzendbureau om geen jonge vrouwen of ouderen te selecteren. Dit zijn tweemaal discriminerende vragen en op die manier kan men discriminatie van belangrijke tussenpersonen onderzoeken. Bij mystery shopping is er ook geen sprake van het vergelijken van quasi-identieke test- en controlepersonen zoals bij praktijktesten.

Mystery shopping stamt uit de commerciële en beleidswereld en is gegroeid in een context van zelfregulering: een bedrijf test zijn eigen processen of laat die testen uitvoeren door een extern, gespecialiseerd consultancybedrijf. Doorgaans worden andere processen dan discriminerend gedrag getest. Het postbedrijf Bpost stuurt bijvoorbeeld mystery shoppers op pad om de algemene dienstverlening van haar postpunten te controleren (Bpost, 2012). De Nationale Loterij controleert haar verkooppunten via minderjarige mystery shoppers op de verkoop van loterijspelen aan minderjarigen.¹ Het bedrijf Carglass bewaakt de dienstverlening van haar medewerkers via mystery shopping.² De methode van mystery shopping kan echter ook zelfregulerend ingezet worden om het voorkomen van discriminatie te testen. Sinds 2011 laat de koepelorganisatie Federgon bijvoorbeeld aangesloten uitzendkantoren opbellen met de vraag om geen 'allochtonen' te sturen (Federgon, 2012, 2013). Achteraf worden de discriminerende uitzendkantoren om uitleg gevraagd en uitgenodigd om een actieplan tegen discriminatie uit te werken. Daarnaast zijn er ook middenveldorganisaties die van de methode van mystery shopping gebruikmaken om discriminatie aan te klagen. Het Minderhedenforum contacteerde bijvoorbeeld dienstenchequebedrijven met de vraag geen poets-hulp van vreemde origine te sturen (Spaas, 2015).

Slechts weinig academische instellingen maken gebruik van mystery shopping. In België en Nederland is ons enkel de masterproef van Evelien Loeters (2011) gekend. Zij onderzocht aan de hand van mystery shopping in welke mate uitzendkantoren in Nederland discrimineren tegen uitzendkrachten van vreemde origine.

Een typologie van discriminatietesten

Een eerste belangrijke onderscheid in onze typologie van discriminatietesten is deze tussen praktijktesten en mystery shopping. Dit onderscheid kan in feite teruggebracht worden tot twee criteria: (1) wordt er al dan niet een discriminerende vraag gesteld en (2) is er al dan niet een test- én controleprofiel? Bij mystery shopping wordt er een discriminerende vraag gesteld en is er geen controleprofiel. Bij praktijktesten wordt er geen discriminerende vraag gesteld en is er wel een controleprofiel. In principe zou men beide types kunnen combineren (bv. men laat een Marokkaanse en Belgische klant

een discriminerende vraag stellen en nadien vergelijkt men of er een ongelijke behandeling was), maar dit is nog nooit gebeurd.

Het derde criterium is of er al dan niet persoonlijk contact is tussen de tester en de geteste organisaties. Bij zowel mystery shopping als praktijktesten kan er wel (telefonisch of in levenden lijve) of geen (per brief of via e-mail) persoonlijk contact zijn. Deze drie verschillpunten resulteren in een typologie van discriminatietesten, gevisualiseerd in Tabel 1.

Tabel 1. Typologie van discriminatietesten.

		Discriminerende vraag?	
		JA	NEE
Persoonlijk contact?	JA	Mystery calls & visits	Situatietesten
	NEE	Mystery mails	Correspondentietesten
		NEE	JA

Test- en controleprofiel?

Wordt er al dan niet een discriminerende vraag gesteld?

Het criterium of er al dan niet een discriminerende vraag gesteld wordt, is om verschillende redenen belangrijk. Het stellen van een discriminerende vraag heeft drie belangrijke gevolgen. Ten eerste bestaat er heel wat discussie over het feit of de vraag om te discrimineren geen verboden vorm van uitlokking zou zijn. Bovendien is niet enkel discrimineren op zich verboden, maar ook de opdracht geven om te discrimineren.³ In tegenstelling tot praktijktesters, zouden mystery shoppers op die manier dus zelf de wet overtreden. Tegelijkertijd kan een vastgoedmakelaar, dienstenchequebedrijf of uitzendkantoor steeds beslissen om niet op de discriminerende vraag in te gaan. Bepaalde vormen van mystery shopping zouden dus mogelijk moeten zijn zonder een verboden vorm van provocatie te zijn. Desalniettemin vormt de juridische discussie onder welke omstandigheden mystery shopping al dan niet een verboden vorm van uitlokking is een belangrijk nadeel van de methode. Een nadeel die de methode van praktijktesten niet heeft.

Ten tweede is mystery shopping maar beperkt toepasbaar. Je kunt enkel mystery shoppen bij dienstverlenende bedrijven waar je een potentiële klant kunt spelen, zoals bijvoorbeeld uitzendkantoren, vastgoedmakelaars, dienstenchequebedrijven of taxibedrijven. Bij bedrijven die hun eigen rekrutering organiseren of bij particuliere verhuurders is het onmogelijk om te vragen of men wil discrimineren. Een belangrijk

voordeel van praktijktesten is daarom dat ze breder toepasbaar zijn: in zowat alle sectoren kun je wel op een of andere manier twee of meerdere kandidaturen uitsuren.

Ten derde is het zo dat ingaan op een discriminerende vraag eigenlijk enkel een intentie tot discrimineren inhoudt. Wat met de vastgoedmakelaar die enkel voor de vorm ingaat op de vraag zodat hij klantvriendelijk overkomt? Indien hij nadien toch beslist om homokoppels of etnische minderheden door te geven als kandidaat-huurders, kunnen we dan besluiten dat we discriminatie hebben vastgesteld? Is ingaan op de discriminerende vraag dan nog voldoende om discriminatie vast te stellen? De intentie hebben om een strafbaar feit te plegen, is niet voldoende. Op zich spreken we enkel van discriminatie indien er sprake is van ongelijke behandeling of uitsluiting, niet wanneer men de intentie heeft om iemand uit te sluiten. Op die manier is mystery shopping wel geschikt om intenties tot discriminatie in kaart te brengen wanneer het gaat om wetenschappelijk onderzoek. Wanneer het gaat om het hard maken van zaken van discriminatie tegen personen of instanties, dan is dit echter moeilijk met mystery shopping.

Is er een test- en controlepersoon?

Bij praktijktesten worden de reacties op de quasi-identieke test- en controlepersoon met elkaar vergeleken, terwijl er bij mystery shopping enkel een testpersoon is die de discriminerende vraag stelt. Het werken met een test- en controlepersoon heeft twee belangrijke implicaties.

Ten eerste kan men met test- en controleprofielen veel beter achterhalen op basis van welk specifiek kenmerk er gediscrimineerd wordt. Door beide profielen zodanig te matchen, kan er immers één specifiek kenmerk geïsoleerd worden. Bij mystery shopping weet men enkel dat een organisatie de intentie heeft om in te gaan op een discriminerende vraag, maar kan men minder zuiver meten waarom men hierop ingaat. Als men bijvoorbeeld vraagt om geen allochtone poetsvrouw te sturen, gaat het dan om discriminatie omwille van de huidskleur, de etnische origine of de religie? Dat is onduidelijk. Het exact kunnen bepalen van de discriminatiegrond is van ondergeschikt belang voor juridische doeleinden (discriminatie is immers discriminatie), maar is wel zeer relevant voor het wetenschappelijke gebruik van discriminatietesten.

Ten tweede dienen praktijktesten herhaaldelijk uitgevoerd te worden, terwijl bij mystery shopping één enkele vaststelling volstaat. Wanneer een bedrijf immers ingaat op een discriminerende vraag, is deze enkele vaststelling reeds voldoende om discriminatie aan te tonen. Een dienstenchequebedrijf kan immers niet 'toevallige selectie' of 'willekeurige ongelijke behandeling' als argumenten aanhalen, wanneer het gewoon ingaat op de vraag om enkel vrouwen van Belgische origine als poetshulp te sturen. Bij praktijktesten daarentegen is er steeds een zekere mate van toeval mogelijk (Heckman & Siegelman, 1993; Ross & Turner, 2005). Hierdoor is het mogelijk dat een testpersoon (bv. van Turkse origine) niet uitgenodigd wordt voor een sollicitatiegesprek of om een huurwoning te bezichtigen en de controlepersoon (bv. van Belgische origine) wel, zonder dat er sprake is van discriminatie. Wanneer een werkgever bijvoorbeeld te veel

kandidaten heeft voor een bepaalde vacature, kan hij beslissen om enkel de eerste tien uit te nodigen. Wanneer de Belgische controlepersoon zich op basis van toeval onder de tien kandidaten bevindt en de Turkse testpersoon niet, dan zou men ten onrechte kunnen stellen dat er gediscrimineerd werd. Een ander voorbeeld is dat een makelaar overstelpt wordt onder de telefoontjes voor een bepaalde huurwoning en vergeet een aantal kandidaten terug te bellen. Wanneer door toeval net de Turkse testpersoon vergeten wordt door de makelaar en de Belgische controlepersoon niet, dan kan men opnieuw onterecht van discriminatie gewag maken. Het omgekeerde is uiteraard ook steeds mogelijk: dat de Turkse testpersoon uitgenodigd wordt en de Belgische controlepersoon niet. Om het element van toeval zo veel mogelijk uit te schakelen, maken onderzoekers daarom gebruik van meerdere praktijktesten. Op die manier kan men willekeurige ongelijke behandeling onderscheiden van een systematisch patroon van discriminatie.

Is er persoonlijk contact of niet?

Bij zowel mystery shopping als praktijktesten zijn er verschillende manieren om deze uit te voeren. Een belangrijk onderscheid is of er persoonlijk contact plaatsvindt of niet. Praktijktesten met persoonlijk contact worden 'situatietesten' genoemd⁴, deze zonder persoonlijke contact 'correspondentietesten'. Mystery shopping met persoonlijk contact noemt men ook 'mystery visits' of 'mystery calls'; mystery shopping zonder persoonlijk contact gebeurt dan weer in de vorm van 'mystery mails'.

Het persoonlijk contact kan zowel telefonisch als face-to-face gebeuren. Zo kunnen sollicitanten op gesprek gaan met een potentiële werkgever. Of er worden dancings bezocht door testgroepen van jongeren. Of er gaat iemand persoonlijk langs bij een vastgoedmakelaar om een huis te verhuren.

Direct persoonlijk contact heeft echter een tweetal nadelen. Ten eerste heb je bij persoonlijk contact minder controle over de omstandigheden van de test en over de verschillen tussen personen (Heckman, 1998; Heckman & Siegelman, 1993). Iemand die mystery shopping uitvoert met persoonlijk contact zou bijvoorbeeld heel hard kunnen aandringen in zijn vraag om te discrimineren. Hierdoor zouden bij mystery shopping hogere cijfers van discriminatie kunnen worden gevonden dan bij onderzoek waar je minder aandringt. Wanneer het gaat om controles door de overheid, zou de invloed van de controleur op het vaststellen van de overtreding dan ook groot zijn. Ook bij praktijktesten zou dit problemen kunnen geven. Stel dat we twee kandidaten uitsuren voor een sollicitatie. Indien één van de twee personen minder vriendelijk, beleefd of verzorgd overkomt dan de andere, dan zou dit gevolgen kunnen hebben voor de vastgestelde discriminatie. Zelfs wanneer je de testpersonen goed opleidt, zijn er toch nog altijd subtiele verschillen tussen twee personen mogelijk (bv. op het vlak van non-verbale communicatie).

Ten tweede is het testen aan de hand van persoonlijk contact zeer tijdsintensief, zowel voor de onderzoekers als voor de verhuurders, uitzendconsulenten of werkge-

vers. Een persoonlijk gesprek dat wordt ingepland met een fictieve verhuurder voor het stellen van een discriminerende vraag zorgt ervoor dat vastgoedmakelaars relatief veel tijd verliezen aan fictieve klanten. Het probleem is nog groter bij praktijktesten. Elke werkgever krijgt dan immers twee fictieve sollicitanten over de vloer voor een tijdrovend sollicitatiegesprek of vergelijkend examen. Zeker wanneer deze praktijktesten een aantal keer herhaald worden, zorgt dit voor een stijgende overlast. Bovendien moet de onderzoeker of overheid dan veel tijd steken in het opleiden en 'vergelijkbaar' maken van paren fictieve kandidaten. Je moet bijgevolg steeds afwegen of de overlast voor de onderzochte personen of instanties niet te groot wordt.

Er is echter ook een manier om via schriftelijk contact de twee methoden uit te voeren. Dan neem je bijvoorbeeld contact op via e-mail of post met bijvoorbeeld een werkgever of verhuurder. Bij mystery shopping stel je dan schriftelijk de vraag om te discrimineren. Bij praktijktesten stellen twee huurkandidaten een quasi-identieke schriftelijke vraag (bv. Mag ik de woning komen bezichtigen?) of sturen ze een quasi-identiek CV en sollicitatiebrief door. De e-mail of brief wordt dan bijvoorbeeld één keer ondertekend met een Vlaamse naam en één keer met een niet-Vlaamse naam om discriminatie op basis van etnische afkomst te onderzoeken. Doordat het contact volledig schriftelijk gebeurt, zijn de resultaten meer gestandaardiseerd. Zo is het vrijwel onmogelijk om extra aan te dringen op discriminatie in het geval van een e-mail met een discriminerende vraag. Een ongelijke behandeling bij praktijktesten is op dat moment niet te wijten aan subtiele verschillen tussen twee fictieve kandidaten, aangezien het om een eenvoudige brief of e-mail gaat. Bovendien is schriftelijk contact ook veel minder tijdsintensief dan persoonlijk contact. Dit is positief voor zowel diegene die de testen uitvoert als de persoon of instantie die getest wordt. De overlast wordt op dat moment immers beperkt tot het beantwoorden van een e-mail, in plaats van het inplannen van een persoonlijk gesprek.

Toch zijn er ook aan schriftelijke testen een aantal nadelen verbonden, specifiek voor elke testmethode. Ten eerste kun je aan de hand van schriftelijke praktijktesten enkel het voorkomen van discriminatie in de eerste fase van het verhuur- of sollicitatieproces nagaan. Je onderzoekt enkel of een kandidaat al dan niet uitgenodigd wordt om een woning te bezichtigen of op een sollicitatiegesprek te komen. In de latere fases, zoals een huisbezoek voor een verhuur of een sollicitatie voor een job, is persoonlijk contact onvermijdelijk. Ten tweede kunnen niet alle discriminatiegronden en alle levensdomeinen met schriftelijke praktijktesten gecontroleerd worden. Terwijl de etnisch-culturele achtergrond of het geslacht van een kandidaat nog duidelijk gemaakt kunnen worden aan de hand van de naam waarmee een brief of e-mail ondertekend wordt, is dit niet of minder mogelijk voor andere discriminatiegronden zoals seksuele geaardheid, religie of leeftijd. Hoe maak je in een brief bijvoorbeeld duidelijk dat je een andere huidskleur hebt? Verhuurders of bedrijven zullen maar raar opkijken wanneer je spontaan je huidskleur vermeldt bij het eerste contact. Schriftelijke testen zijn bovendien ook niet mogelijk in de horeca of in de transportsector. Je stuurt immers geen e-mail met de vraag een dancing al dan niet binnen te mogen of al dan niet een taxi of een bus te mogen nemen. Ten derde heeft het medium ook

een invloed op de graad van discriminatie. Discrimineren bij schriftelijk contact is gemakkelijker bij praktijktesten (Heylen & Van den Broeck, 2015). Het is gemakkelijker, anoniemer en vrijblijvender om iemand geen antwoord te sturen wanneer je een e-mail ontvangt van een kandidaat die je wilt uitsluiten dan wanneer je deze aan de telefoon hebt of in levenden lijve ontmoet. Onderzoek of controles aan de hand van schriftelijke praktijktesten zullen daarom meer discriminatie vaststellen dan telefonische praktijktesten.

Het gebruik van praktijktesten als beleidsinstrument

In Europa werden praktijktesten tot nu toe voornamelijk uitgevoerd om discriminatie te onderzoeken en nauwelijks om discriminatie effectief terug te dringen. Recent komt er hier verandering in en gaan steeds meer stemmen op om praktijktesten ook als beleidsinstrument in te zetten in de strijd tegen discriminatie. In een aantal steden werden hiervoor reeds concrete stappen ondernomen. Het onderzoek naar praktijktesten als beleidsinstrument staat echter nog in zijn kinderschoenen.

Werken praktijktesten in de strijd tegen discriminatie?

Een eerste evidente vraag is uiteraard of praktijktesten überhaupt wel effectief zijn om discriminatie terug te dringen. Dit wordt doorgaans als waar aangenomen, maar onderzoek hiernaar is schaars. Wanneer we naar de Verenigde Staten kijken, dan stellen we vast dat de discriminatie van Afro- en Hispanic-Amerikanen op de woningmarkt de afgelopen decennia zeer fors gedaald is (Ross & Turner, 2005; Turner et al., 2013). Het zou echter te voorbarig zijn om deze daling enkel toe te schrijven aan de inzet van praktijktesten. Er zijn verschillende alternatieve verklaringen mogelijk, gaande van sensibilisering bij en een betere training van vastgoedmakelaars tot de groeiende sociaal-economische mogelijkheden van de Afro- en Hispanic-Amerikaanse middenklasse. Een studie van Ross en Galster (2007) in de VS suggereert echter dat in steden waar er praktijktesten uitgevoerd werden én waar betrapte vastgoedmakelaars financieel beboet werden, de discriminatie op de woningmarkt vervolgens afnam. De vraag hierbij is of makelaars ook minder zouden discrimineren indien ze enkel getest werden door middel van praktijktesten en niet beboet? Anders gesteld, werken praktijktesten zonder sanctionering? De achterliggende idee hierbij is dat mensen hun gedrag aanpassen conform de sociale of wettelijke verwachtingen, wanneer ze weten dat ze gecontroleerd of geobserveerd worden. Deze idee stamt uit de Hawthorne- en sociale wenselijkheidstheorieën (Cialdini & Goldstein, 2004; Jones, 1991; Levitt & List, 2011).

Deze vraag stond centraal in een quasi-experimentele studie naar etnische discriminatie op de Gentse huisvestingsmarkt (Van der Bracht & Verhaeghe, 2017). De onderzoekers voerden er twee golven van correspondentietesten uit. In de eerste golf waren de particuliere verhuurders en makelaars niet op de hoogte van de

praktijktesten, in de tweede golf wel. Tussen beide golven organiseerden de onderzoekers een grote bekendmakingscampagne. De resultaten van de eerste golf, waaruit grootschalige discriminatie bleek, werden uitvoerig in de pers besproken én er werd een tweede golf van praktijktesten aangekondigd. Bovendien kregen alle vastgoedmakelaars tijdens de campagne een persoonlijke brief, ondertekend door de burgemeester en het schepencollege van Stad Gent. In de brief werd nogmaals de nieuwe golf van praktijktesten aangekondigd, maar met de mededeling dat ze geen juridische of financiële consequenties voor de betrokken makelaars zouden hebben. De particuliere verhuurders konden helaas niet persoonlijk gecontacteerd worden, omwille van het gebrek aan een officiële databank van particuliere verhuurders. Tussen de eerste en tweede golf daalde de ongelijke behandeling van etnische minderheden aanzienlijk bij vastgoedmakelaars, maar niet bij particuliere verhuurders. Deze resultaten suggereren dat het controleren van makelaars door middel van praktijktesten zonder evenwel te sanctioneren, al tot een gedragsverandering kan leiden. De afwezigheid van een significante daling in discriminatie bij particuliere verhuurders suggereert bovendien dat een persoonlijke aanpak (bv. via een persoonlijke brief) een belangrijke factor is om tot een meer sociaal wenselijke gedragswijziging te komen.

Er blijven evenwel nog veel onduidelijkheden. Eerst en vooral kan men zich afvragen in welke mate men de resultaten van het experiment op de Gentse woningmarkt kan veralgemenen. In een soortgelijk experiment onderzochten Carlsson en Rooth (2009) het effect van praktijktesten op het discriminerende gedrag van werkgevers op de Zweedse arbeidsmarkt. In de tweede golf van het experiment waren de werkgevers van de praktijktesten op de hoogte na een grootschalige bekendmaking in de media (er was echter geen persoonlijke communicatie). De onderzoekers vonden geen verandering in het discriminerende gedrag van werkgevers tussen de eerste en tweede golf. Het is onduidelijk of dit te wijten is aan de afwezigheid van een persoonlijke aanpak, de sector van de arbeidsmarkt of andere contextuele factoren. Ten tweede weten we nog niet of het effect van praktijktesten zonder sanctionering duurzaam is of niet. In het Gentse experiment stelde men een gedragsverandering bij makelaars vast tijdens de tweede golf van correspondentietesten (Van der Bracht & Verhaeghe, 2017). De vraag is of de makelaars blijvend ophouden met discrimineren zodra ze niet meer gecontroleerd worden. Andere studies naar sociaal wenselijk gedrag (bv. handhygiëne in ziekenhuizen of spaarzaam energieverbruik bij gezinnen) suggereren evenwel een eerder kortstondig effect (Schwartz, Fischhof, Krishnamurti & Sowell, 2013; Srigley, Furness, Baker & Gardam, 2014). Zodra de controle wegvalt, hervalt men in het oude, ongewenste gedrag. Beleidsmatig zou dit betekenen dat indien praktijktesten zonder sanctionering al zouden werken tegen discriminatie, men blijvend of toch op regelmatige tijdstippen de praktijktesten zouden moeten uitvoeren.

Zijn praktijktesten haalbaar om individuele organisaties te testen?

Een belangrijk onderscheid tussen mystery shopping en praktijktesten is dat bij mystery shopping één enkele vaststelling volstaat, terwijl praktijktesten herhaaldelijk uitgevoerd moeten worden. Meerdere praktijktesten zijn noodzakelijk om een toevallige ongelijke behandeling te kunnen onderscheiden van een systematisch patroon van discriminatie. Meerdere praktijktesten zijn doorgaans geen probleem wanneer men het voorkomen van discriminatie in een hele sector wil testen voor onderzoeksdoeleinden. De praktijktesten worden dan over meerdere organisaties, bedrijven of personen verdeeld. Voor juridische doeleinden wil men evenwel geen uitspraken doen over het voorkomen van discriminatie in een bepaalde sector, maar wel over het potentieel discriminerende gedrag van een individuele actor. Dit betekent dat men die individuele organisatie of persoon meermaals moet kunnen testen door middel van praktijktesten.

Vanuit beleidsoogpunt is een belangrijke vraag bijgevolg hoeveel praktijktesten per individuele actor noodzakelijk zijn? Een meta-analyse naar etnische discriminatie toonde aan dat het aantal praktijktesten op sectorniveau varieert van 92 tot 3.153, met een mediaan van 759 testen op de woningmarkt, en van 240 naar 1.552, met een mediaan van 754 testen op de arbeidsmarkt (Rich, 2014, eigen berekeningen). Een dergelijk hoog aantal praktijktesten is zowel onhaalbaar als onwenselijk op individueel niveau. Onhaalbaar omdat een makelaar of werkgever simpelweg te weinig huuradvertenties of openstaande vacatures heeft om te kunnen testen. Onwenselijk omdat de overlast van het testen te hoog is om door een individuele actor gedragen te kunnen worden. Een veel lager aantal praktijktesten per actor is dus nodig. Uit een longitudinale studie van Verhaeghe en Van der Bracht (2016) bij 114 vastgoedmakelaars in België blijkt dat er minstens tien testen per makelaar nodig zijn om discriminerend gedrag te kunnen vaststellen. Om het significantieniveau te bepalen, gebruikten ze hierbij de McNemar mid-p-test. Deze test is ideaal voor kleine steekproeven met gematchte data (Fagerland, Lydersen & Laake, 2013; Vuolo, Uggen & Lageson, 2015).

Het benodigd aantal praktijktesten per actor hangt echter af van drie factoren (Verhaeghe & Van der Bracht, 2016). Ten eerste is het gemakkelijker een significant patroon van discriminatie vast te stellen bij organisaties die consequent discrimineren dan bij organisaties die slechts af en toe discrimineren. Bij die laatste actoren zullen er veel meer dan tien praktijktesten nodig zijn om de discriminatie te kunnen vaststellen, terwijl bij de eerste actoren reeds vijf praktijktesten kunnen volstaan. Dit is niet enkel een methodologische afweging, maar houdt ook een beleidskeuze in. Wil men enkel de hardleerse discriminerende organisaties controleren en sanctioneren of wil men ook de organisaties die slechts sporadisch discrimineren aanpakken?

Ten tweede is er de mate van zekerheid die men over het voorkomen van discriminatie wil hebben. Of anders gesteld: in hoeverre wil men de toevalsfactor (statistisch uitgedrukt in de p-waarde) minimaliseren? In veel landen volstaat al het 'vermoeden van discriminatie' om juridische stappen te kunnen zetten (Rorive, 2009). Wanneer

men deze piste volgt, volstaan er minder strenge p-waarden en moeten er bijgevolg minder praktijktesten per actor uitgevoerd worden. Ook dit aspect is een beleidskeuze: hoeveel zekerheid wil men hebben vooraleer men een organisatie op de vingers tikt? Men kan dit eventueel laten afhangen van de gevolgen. Een licht vermoeden van discriminatie kan bijvoorbeeld volstaan om een organisatie uit te nodigen om een diversiteitstraining te volgen, terwijl een sterk significant patroon van discriminatie vereist is om een organisatie financieel te beboeten. Dit is uiteraard een continuüm en het is aan beleidsmakers om hierin een beslissing te nemen.

Een laatste factor is de non-respons. Non-respons bij praktijktesten verwijst naar de situatie waarbij zowel de testpersoon als de controlepersoon geen reactie van een makelaar, werkgever of andere organisatie krijgt. Hoe hoger de non-respons, hoe meer praktijktesten er bij een actor noodzakelijk zijn om significante uitspraken over diens potentieel discriminerend gedrag te kunnen doen. Een populaire werkgever bijvoorbeeld die om praktische redenen slechts heel weinig sollicitanten kan uitnodigen, is dus moeilijker op discriminatie te testen dan een werkgever die een vacature voor een knelpuntberoep slechts moeizaam ingevuld krijgt en bijgevolg bijna alle kandidaten uitnodigt. Uit voorgaand onderzoek blijkt dat discriminatie net het meeste voorkomt wanneer de vraag groot is en het aanbod beperkt, zoals bij goedkope huurwoningen (Carpusor & Loges, 2006; Ahmed & Hammarstedt, 2008; Bosch, Carnero & Farré, 2010; Hanson & Hawley, 2011) of niet-knelpuntberoepen (Baert et al., 2015). Met andere woorden: hoe groter de vraag, hoe meer discriminatie, maar ook hoe hoger de non-respons en bijgevolg hoe moeilijker op discriminatie te testen.

Zijn praktijktesten ethisch verantwoord?

In het maatschappelijke debat bestaat er veel commotie over de ethiek van praktijktesten. Ook al is er geen sprake van uitlokking zoals bij de techniek van mystery shopping, toch zijn er een aantal ethische kwesties die uitgeklaard moeten worden. Tegenstanders van praktijktesten halen drie ethische problemen aan.

Het eerste bezwaar is dat de geteste organisaties deelnemen aan een onderzoek waarvan ze niet op de hoogte zijn en waarvoor ze bijgevolg ook geen toestemming gegeven hebben. Dit argument geldt zowel voor praktijktesten als academische onderzoeksmethode als voor praktijktesten als beleidsinstrument. Zeker in de onderzoeksweld is het verkrijgen van de zogenaamde '*informed consent*' een belangrijk ethisch principe. Verschillende onderzoekers verdedigen niettemin het gebruik van praktijktesten met het argument dat het de enige manier is om betrouwbare informatie over discriminerend gedrag te verkrijgen (Banton, 1997; Bovenkerk, 1992; Riach & Rich, 2004; Zschirnt, 2016b). Werkgevers, makelaars of andere organisaties informeren over de praktijktesten zou het onderzoek minder betrouwbaar en valide maken. Dit tegenargument wordt doorgaans ook door ethische commissies gevolgd (Zschirnt, 2016b). Bovendien stelt Bovenkerk (1992) dat discriminatie bij bijvoorbeeld een aanwerving niet langer om een private aangelegenheid gaat, omdat de maatschappelijke positie

van grote groepen van mensen erdoor getroffen wordt. Men zou dezelfde tegenargumenten kunnen maken bij het gebruik van praktijktesten als beleidsinstrument, maar veel juridisch of sociologisch onderzoek is hiernaar nog niet verricht.

Een tweede kritiek op praktijktesten is dat de betrokken actoren misleid worden. Men doet zich bijvoorbeeld voor als een potentiële kandidaat voor een job of voor een huurwoning, zonder evenwel een echte kandidaat te zijn. De onderzoekers Riach en Rich (2004) erkennen dit probleem, maar stellen dat deze vorm van misleiding een kleine prijs is dat men moet betalen voor de strijd tegen discriminatie. Het nadeel weegt met andere woorden niet op tegen het voordeel.

Een laatste argument tegen praktijktesten is dat onderzoeksobjecten geen schade mogen lijden van het onderzoek. Deze schade kan minstens twee vormen aannemen. Allereerst mag de overlast van het testen zelf niet disproportioneel hoog liggen. In iedere praktijktest moet een organisatie immers kostbare tijd steken. Bij academisch onderzoek gaat men een organisatie daarom doorgaans slechts één keer testen en nooit meer dan twee of drie keer (Zschirnt, 2016b). Dit is evenwel volledig anders bij het gebruik van praktijktesten als beleidsinstrument. Hier wil men geen uitspraken doen over een hele sector, maar wel over een individuele actor, waardoor meermaals testen per organisatie noodzakelijk is. Mogelijke uitwegen zijn hier om zeer gericht te gaan testen (bv. na een specifieke klacht over een makelaar of interimkantoor), zich tevreden te stellen met het aantonen van slechts een vermoeden van discriminatie waardoor er minder praktijktesten per organisatie nodig zijn, en de praktijktesten zeer beknopt te houden (bv. enkel nagaan of iemand door een werkgever uitgenodigd wordt in plaats van effectief op sollicitatiegesprek te gaan).

Een tweede vorm van schade heeft betrekking op de gevolgen van de praktijktesten voor de geteste organisatie. Hier verschillen de praktijktesten als onderzoeksmethode en deze als beleidsinstrument fundamenteel van elkaar. Wanneer men discriminatie louter wil onderzoeken, dan moeten de onderzoekers er alles aan doen om eventuele nadelige gevolgen voor de geteste organisaties te vermijden (Pager, 2007; Zschirnt, 2016b). Het bewaren van de anonimiteit van de onderzoeksobjecten is hiervoor een gouden regel. Bij praktijktesten als beleidsinstrument is het doel echter het terugdringen van discriminatie. Men streeft naar een gedragsverandering bij de onderzoeksobjecten. Naargelang de beleidskeuze kunnen discriminerende organisaties al dan niet juridisch vervolgd worden. Mogelijke sancties kunnen variëren van het verplicht volgen van een diversiteitstraining, het betalen van een boete of schadevergoeding tot de intrekking van een licentie. Zelfs wanneer men louter inzet op een controlesysteem van praktijktesten zonder evenwel te sanctioneren, kan een werkgever of makelaar dit als nadelig beschouwen. Hij wordt zo onder druk gezet om zijn gedrag te veranderen. In deze gevallen moeten beleidsmakers weer de afweging maken of de mogelijke nadelige gevolgen voor individuele organisaties opwegen tegen het maatschappelijke doel van meer gelijke kansen.

Conclusie

Discriminatie is een hardnekkig probleem van onze samenleving. Etnisch-culturele minderheden, vrouwen, ouderen, holebi's en tal van andere sociale groepen worden structureel en ongeoorloofd benadeeld. Tegen deze achtergrond gaan er steeds meer stemmen op om praktijktesten in te voeren. In de wetenschappelijke wereld bestaat er reeds een lange traditie van het gebruik van praktijktesten als onderzoeksmethode om discriminatie te bestuderen. Het onderzoek naar praktijktesten als beleidsinstrument staat evenwel nog in haar kinderschoenen.

Eerst en vooral is het belangrijk dat er in het beleidsonderzoek en het maatschappelijke debat meer duidelijkheid komt over wat de techniek van praktijktesten precies inhoudt en wat de voor- en nadelen zijn. Anders schuilt het gevaar dat academici, opiniemakers en politici naast elkaar praten. In dit overzichtsartikel hebben we daarom een typologie van discriminatietesten voorgesteld. We maakten een conceptueel onderscheid tussen praktijktesten en mystery shopping op basis van twee criteria: (1) het al dan niet stellen van een discriminerende vraag en (2) het al dan niet werken met een test- en controlepersoon. Door dit onderscheid vervalft reeds de vaak gehoorde kritiek op praktijktesten dat ze een verboden vorm van uitlokking zijn. Vervolgens splitsten we de technieken van praktijktesten en mystery shopping verder op naargelang er al dan niet persoonlijk contact is. Dit resulteerde in vier types van discriminatietesten: situatietesten, correspondentietesten, mystery calls/visits en mystery mails. Deze typologie is zeker niet algemeen aanvaard, maar ze kan toekomstig onderzoek conceptueel sturen.

In het tweede deel van dit overzichtsartikel bespraken we de stand van zaken van het onderzoek naar het gebruik van praktijktesten om discriminatie terug te dringen. Er is beperkte wetenschappelijk evidentie dat praktijktesten in bepaalde omstandigheden en onder bepaalde voorwaarden effectief, haalbaar en ethisch verantwoord zijn. Wij zijn daarom voorzichtige voorstanders van het gebruik van praktijktesten als beleidsinstrument tegen discriminatie. Desalniettemin is er nog veel meer beleidsonderzoek vereist. Nog los van de haalbaarheid en ethische uitvoering van praktijktesten is de belangrijkste onderzoeksvraag deze naar hun effectiviteit. In welke situaties dringen praktijktesten discriminatie terug en hoe duurzaam zijn hun effecten?

Belangrijk is evenwel dat voorstanders van praktijktesten beseffen dat de techniek zeker geen wondermiddel is. De strijd tegen discriminatie vergt een totaalaanpak, waarbij praktijktesten gecombineerd worden met tal van andere acties, gaande van sensibilisering rond vooroordelen tot sanctionering en het vergroten van de machtsmiddelen van achtergestelde groepen.

Noten

1. <http://www.nationale-loterij.be/nl/over-ons/missie-en-opdracht/verantwoord-spel/minderjarigen>

2. <http://www.mobius.eu/case/carglass/>
3. Wet van 10 mei 2007 ter bestrijding van bepaalde vormen van discriminatie.
4. Er zijn hier weer verschillen tussen de Amerikaanse en Europese context. In de VS gebruikt men de term 'audittesten' ('*audit tests*') voor praktijktesten met persoonlijk contact, terwijl in Europa men de term 'situatietesten' ('*situation tests*' of '*tests de situation*') hiervoor gebruikt (Riach & Rich, 2002).

Bibliografie

- Ahmed, A. & Hammarstedt, M. (2008). Discrimination in the rental housing market: a field experiment on the internet. *Journal of Urban Economics*, 64, 362-372.
- Baert, S., Cockx, B., Gheyle, N. & Vandamme, C. (2015). Is there less discrimination in occupations where recruitment is difficult? *Industrial & Labor Relations Review*, 68, 467-500.
- Baert, S. & Omeij, E. (2015). Hiring discrimination against pro-union applicants: the role of union density and firm size. *Economist*, 163, 263-280.
- Baert, S. & Verhofstadt, E. (2015). Labour market discrimination against former juvenile delinquents: evidence from a field experiment. *Applied Economics*, 47, 1061-1072.
- Baert, S., De Pauw, A-S. & Deschacht, N. (2016). Do employer preferences contribute to sticky floors. *ILR Review*, DOI:10.1177/0019793915625213.
- Banton, M. (1997). The ethics of practice-testing. *Journal of Ethnic and Migration Studies*, 23, 413-420.
- Bertrand, M. & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94, 991-1013.
- Bolt, D. (2005). From blindness to visual impairment: terminological typology and the social model of disability. *Disability & Society*, 20, 539-552.
- Bosch, M., Carnero, A. & Farré, L. (2010). Information and discrimination in the rental housing market: evidence from a field experiment. *Regional Science and Urban Economics*, 40, 11-19.
- Bovenkerk, F. (1992). *A Manual for International Comparative Research on Discrimination on the Grounds of 'Race' and 'Ethnic origin'*. Geneva: International Labour Organization.
- Bpost (2012). *Jaarverslag 2012*. Brussel: Bpost.
- Capéau, B., Eeman, L., Groenez, S. & Lamberts, M. (2012). *Wie heeft voorrang: jonge Turken of prille grijsaards? Een experimenteel onderzoek naar discriminatie op basis van persoonkenmerken bij de eerste selectie van sollicitanten – technisch verslag*. Leuven: HIVA.
- Carlsson, M. & Eriksson, S. (2015). Ethnic discrimination in the London market for shared housing. *Journal of ethnic and migration studies*, 41, 1276-1301.
- Carlsson, M. & Rooth, D. O. (2009). *The impact of being monitored on discriminatory behavior among employers: evidence from a natural experiment*. IZA Discussion Paper No. 3972. Bonn: The Institute for the Study of Labor.
- Carpusor, A. G. & Loges, W. E. (2006). Rental discrimination and ethnicity in names. *Journal of applied social psychology*, 36, 934-952.
- Ceobanu, A. & Escandell, X. (2010). Comparative analyses of public attitudes toward immigrants and immigration using multinational survey data: a review of theories and research. *Annual Review of Sociology*, 36, 309-328.

- Cialdini R. & Goldstein, N. (2004). Social influence: Compliance and Conformity. *Annual Review of Psychology*, 55, 591-621.
- Coenders, M. & Scheepers, P. (2008). Changes in resistance to the social integration of foreigners in Germany 1980-2000: individual and contextual determinants. *Journal of Ethnic and Migration Studies*, 34, 1-26.
- Coenders, M., Lubbers, M., Scheepers, P. & Verkuyten, M. (2008). More than two decades of changing ethnic attitudes in the Netherlands. *Journal of Social Issues*, 64, 269-285.
- Daniel, W. (1968). *Racial Discrimination in England*. Middlesex: Penguin Books.
- Essed, P. (1991). *Understanding everyday racism: An interdisciplinary theory*. Newbury Park, CA: Sage.
- Fagerland, M., Lydersen, S. & Laake, P. (2013). The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. *BMC Medical Research Methodology*, 13, 1-8.
- Federgon (2012). *Jaarverslag 2011*. Brussel: Federgon.
- Federgon (2013). *Jaarverslag 2012*. Brussel: Federgon.
- Fix, M. & Turner, M.A. (1998). *Report Card on Discrimination in America: the Role of Testing*. Washington, DC: Urban Institute Press.
- Ford, R. (2008). Is racial prejudice declining in Britain? *The British Journal of Sociology*, 59, 609-636.
- Hanson, A. & Hawley, Z. (2011). Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities. *Journal of Urban Economics*, 70, 99-114.
- Heckman, J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, 12, 101-116.
- Heckman, J. & Siegelman, P. (1993). The Urban Institute audit studies: Their methods and findings. In M. Fix, R.J. Struyk (Eds.), *Clear and Convincing Evidence: Measurement of Discrimination in America* (pp. 187-258). Washington, DC: Urban Institute Press.
- Heylen, K. & Van den Broeck, K. (2015). Discrimination and selection in the Belgian private rental market. *Housing Studies*. DOI: 10.1080/02673037.2015.1070798.
- Jones, S. (1991). Was There a Hawthorne Effect? *American Journal of Sociology*, 98, 451-485.
- Jowell, R. & Prescott-Clarke, P. (1970). Racial discrimination and white-collar workers in Britain. *Race & Class*, 11, 397-417.
- Levitt, S. & List, J. (2011). Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. *American Economic Journal of Applied Economy*, 3, 224-238.
- Loeters, E. (2011). *De klant is koning. Een onderzoek naar het honoreren van discriminerende verzoeken van werkgevers door intercedenten van uitzendbureaus in Nederland*. Vrije Universiteit Amsterdam: Masterthesis.
- Meuleman, B., Davidov, E. & Billiet, J. (2009). Changing attitudes toward immigration in Europe, 2002-2007: a dynamic group conflict theory approach. *Social Science Research*, 38, 352-365.
- Özdil, Z. (2015). *Nederland mijn vaderland*. Amsterdam: De Bezige Bij.
- Pager, D. (2007). The use of field experiments for studies of employment discrimination: contributions, critiques and directions for the future. *Annals of the American Academy of Political and Social Science*, 609, 104-133.
- Pager, D. & Shepherd, H. (2008). The sociology of discrimination: racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, 34, 181-209.
- Pager, D. & Western, B. (2012). Identifying discrimination at work: the use of field experiments. *Journal of Social Issues*, 68, 221-237.

- Riach, P. A. & Rich, J. (2002). Field experiments of discrimination in the market place. *The Economic Journal*, 112, 480-518.
- Riach, P. A. & Rich, J. (2004). Deceptive field experiments of discrimination: Are they ethical? *Kyklos*, 57, 457-470.
- Rich, J. (2014). *What do field experiments of discrimination in markets tell us? A meta-Analysis of studies conducted since 2000. IZA Discussion Paper No. 8584*. Bonn: The Institute for the Study of Labor.
- Risman, B. (2004). Gender as a social structure. Theory wrestling with activism. *Gender & Society*, 18, 429-450.
- Rorive, I. (2009). *Proving Discrimination Cases: the Role of Situation Testing*. Brussels: Migration Policy Group & Centre For Equal Rights.
- Ross, S. & Galster, G. (2007). Fair housing enforcement and changes in discrimination between 1989 and 2000. An exploratory study. In J. Goering (Ed.), *Fragile rights within cities. Government, housing and fairness*, 177-202. Plymouth: Rowman & Littlefield Publishers.
- Ross, S. & Turner, M. (2005). Housing discrimination in metropolitan America: explaining changes between 1989 and 2000. *Social Problems*, 52, 152-180.
- Semyonov, M. & Gorodzeisky, A. (2006). The rise of anti-foreigner sentiment in European societies, 1988-2000. *American Sociological Review*, 71, 426-449.
- Spaas, N. (2015). *Dienstencheques: subsidiëren om te discrimineren?* Brussel: Minderhedenforum.
- Srigley, J., Furness, C., Baker, G. & Gardam, M. (2014). Quantification of the Hawthorne effect in hand hygiene compliance monitoring using an electronic monitoring system: a retrospective cohort study. *BMJ Qual Saf*, 23, 974-80.
- Schwartz, D., Fischhoff, B., Krishnamurti, T. & Sowell, F. (2013). The Hawthorne effect and energy awareness. *Proceedings of the National Academy of Science of the United States of America*, 110, 15242-15246.
- Turner, M. & Ross, S. (2005). How Racial Discrimination Affects the Search for Housing. In X. S. Briggs (Ed.), *The Geography of Opportunity: Race and Housing Choice in Metropolitan America* (81-100). Washington, DC: Brookings Institution Press.
- Turner, M. A., Santos, R., Levy, D., Wissoker, D., Aranda, C. & Pitingolo, R. (2013). *Housing Discrimination against Racial and Ethnic Minorities 2012*. Washington, DC: US Department of Housing and Urban Development.
- Van der Bracht, K., Coenen, A. & Van de Putte, B. (2015). The not-in-my-property syndrome: the occurrence of ethnic discrimination in the rental housing market in Belgium. *Journal of Ethnic and Migration Studies*, 41, 158-175.
- Van der Bracht, K. & Verhaeghe, P. P. (2016). *Gelijke toegang tot huisvesting voor elke Gentenaar: onderzoeksrapport*. Gent: Vakgroep Sociologie.
- Van der Bracht, K. & Verhaeghe, P. P. (2017). Can racism be reduced by inducing compliance? A quasi-experiment in rental housing discrimination. *Revised and resubmit to Plos One*.
- Verhaeghe, P. P., Van der Bracht, K. & Van de Putte, B. (2016). Discrimination of tenants with a visual impairment on the housing market: empirical evidence from correspondence tests. *Disability and Health Journal*, 9, 234-238.
- Verhaeghe, P. P. & Van der Bracht, K. (2016). *How many correspondence tests are enough to detect discrimination among single agents? A longitudinal study on the Belgian real estate market. Working paper*. Gent: Vakgroep Sociologie.
- Vuolo, M., Uggen, C. & Lageson, S. (2015). Statistical power in experimental audit studies: cautions and calculations for matched tests with nominal outcomes. *Sociological Methods and Research*. DOI: 10.1177/0049124115570066.

- Yinger, J. (1998). Testing for discrimination in housing and related markets. In M. Fix, M. Austin Turner (Eds.), *A national report card on discrimination in America: the role of testing*. Washington, DC: The Urban Institute.
- Zschirnt, E. (2016a). Measuring hiring discrimination – a history of field experiments in discrimination research. *Working paper series #7*, NCCR On the Move.
- Zschirnt, E. (2016b). Revisiting ethics in correspondence testing. *Working paper series #8*, NCCR On the Move.
- Zschirnt, E. & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990-2015. *Journal of Ethnic and Migration Studies*, 42, 1115-1134.

Abstract

During the last years, situation tests and mystery shopping got much attention in public debates. Academics uncovered structural discrimination on both housing and labour markets and politicians and civil society movements made a strong plea for situation tests and mystery shopping as means to tackle discriminatory companies and realtors. This review article examines to which extent situation and correspondence tests can be used as both a scientific method and a policy instrument in the fight against discrimination. First, we construct on the basis of three criteria a typology of techniques to test discrimination: situation tests, correspondence tests, mystery calls/visits and mystery mails. For each test, we discuss briefly their history, applications and ad- and disadvantages. This typology is not yet widely accepted, but can conceptually feed future research. Second, we make a *status quaestionis* of the research about situation and correspondence tests a policy instruments. Research is scarce, but suggests that in some contexts and under some conditions situation and correspondence tests are effective, realistic and ethical to perform. Therefore, we are in favour of using situation and correspondence tests as a policy instrument to tackle discrimination.

Keywords

situation tests, correspondence tests, discrimination, housing market, labour market, policy, methodology
