

Selectie voor Diversiteit: Naar een inclusieve aanpak

Eva Deraus

Samenvatting

Mensen maken organisaties: het selecteren van goede werknemers is essentieel voor het garanderen van goed functionerende organisaties. De arbeidsmarkt is de laatste decennia meer divers geworden en een belangrijk vraag is of selectiemethodes aangepast zijn aan de diverse doelgroepen op onze arbeidsmarkt. Vanuit twee dominante perspectieven op selectie bekijken we eerst hoe we op een meer inclusieve manier kunnen selecteren via redelijke testaanpassingen. We bespreken vervolgens de belangrijkste selectietests en uitdagingen die redelijke aanpassingen stellen aan test validiteit en test fairness en schetsen een experimenteel onderzoek naar het effect van redelijke aanpassingen op testprestaties bij personen met/zonder een visuele beperking dat uitgevoerd werd bij Selor. We besluiten met een alternatief model voor een meer inclusief selectiebeleid.

Auteur noot. Eva Deraus is werkzaam aan de Vakgroep personeelsmanagement, Arbeids- en Organisatiepsychologie van de Faculteit Psychologie en Pedagogische Wetenschappen te Gent. Zij dankt Selor - FOD BOSA en in het bijzonder Els Daems (werkgroep Diversiteit Selor) en Bieke Van den Bon (werkgroep Redelijke Aanpassingen Selor) voor de geboden mogelijkheid te kunnen rapporteren over de experimentele studie naar redelijke aanpassingen van selectietests voor personen met een visuele beperking. De auteur maakte destijds deel uit van de werkgroep die zich bezig hield met de evaluatie van redelijke aanpassingen aan selectietest voor personen met een visuele beperking.

Op Zoek naar Talent

De maatschappij is de laatste decennia erg divers geworden: etnisch-culturele meerderheden en minderheden, mannen en vrouwen, jongeren en ouderen, mensen zonder en met (fysieke/psychische) beperkingen zoeken hun weg naar de arbeidsmarkt. Dat is goed nieuws. Talent op de arbeidsmarkt is immers schaars. Hoe groter en diverser de groep werkzoekenden wordt, hoe meer kans werkgevers hebben om talentvolle werknemers te vinden en hoe meer ‘verschillende’ talentvolle mensen aan het werk zijn. Openstaan voor diversiteit leidt zo tot een ‘win-win’ situatie voor iedereen. Tot zover de theorie, maar volgt ook de praktijk? Paradoxaal genoeg krijgen niet alle werkzoekenden evenveel kansen op de arbeidsmarkt. Zo zijn mensen met een vreemde origine, mensen die ouder zijn en mensen met een beperking nog steeds ondervertegenwoordigd op de arbeidsmarkt in vergelijking met mensen met een niet-vreemde origine, die jong zijn en geen beperking hebben (Neumark, Burn, & Button 2016; Vornholt et al. 2017; Zschirnt & Ruedin 2016). Uiteraard is de problematiek van gelijke arbeidsmarktkansen complex en multifactorieel. In dit hoofdstuk zoomen we in op één specifiek aspect, nl. selectieprocedures in organisaties. Het selecteren van goede werknemers is immers essentieel voor het garanderen van goed functionerende organisaties. Maar zijn deze selectieprocedures ook aangepast aan de diverse en specifieke noden van alle werkzoekenden op onze arbeidsmarkt ?

Selectieprocedures bepalen wie de organisatie binnenkomt (‘select in’) en wie de organisatie niet binnenkomt (‘select out’). Het doel van personeelsselectie is immers een onderscheid te maken tussen sollicitanten (of kandidaten) op basis van hun potentieel en geschiktheid voor de functie waarvoor ze solliciteren. Kortom, als verschillende kandidaten het later op de werkplek even goed zouden doen in een bepaalde functie, dan moeten ze een even grote kans hebben op een positief advies tijdens de selectieprocedure. Een werkgever, HR manager, selectieconsultant of selectiepsycholoog is dus bij uitstek iemand die discrimineert. Want selecteren is per definitie discrimineren of een onderscheid maken tussen mensen. De term ‘discriminatie’ heeft maatschappelijk een negatieve connotatie maar met discriminatie in de context van personeelsselectie is dus niets mis zolang deze discriminatie terecht of functie-relevant is. Alleen kan discriminatie binnen een selectiecontext ook onterecht zijn wanneer een kandidaat (on)bewust beoordeeld wordt op kenmerken die niet functie-relevant zijn (door vooroordelen of misconcepties; Fraser et al. 2010) of wanneer de testprocedure onvoldoende rekening houdt met niet-functierelevante kenmerken en kwetsbare situaties

waarin kandidaten zich bevinden. Een recente meta analyse van 42 correspondentietests¹ in 18 verschillende landen (periode 1990-2015) toont, bijvoorbeeld, hoe kandidaten uit etnisch-culturele minderhedengroepen in vergelijking met even sterk gekwalificeerde kandidaten uit etnisch-culturele meerderheden groepen tot 50% meer sollicitatiebrieven moeten opsturen naar potentiële werkgevers om evenveel kans te maken op een uitnodiging voor een selectiegesprek (Zschirnt & Ruedin 2016). Selectiepraktijken kunnen kandidaten dus ook onterecht een redelijke kans op werk ontnemen. Onterechte discriminatie in een selectiecontext is niet toelaatbaar vanuit sociaal-maatschappelijk, economisch als ethisch perspectief. In sommige situaties is onterechte discriminatie zelfs wettelijk strafbaar.

In de wetenschappelijke literatuur wordt vooral aandacht besteed aan (terechte/onterechte) discriminatie van sollicitanten uit de drie grootste diversiteitsgroepen, nl. etniciteit (etnisch-culturele minderheden vs. meerderheid), leeftijd (oudere vs. jongere kandidaten), en geslacht (vrouwen vs. mannen). Thema's die bestudeerd worden, zijn de mate van (on)terechte discriminatie, determinanten van (on)terechte discriminatie en manieren om onterechte discriminatie te vermijden (Deros & Ryan 2014). Opvallend, een systematische literatuurstudie (via 'Web of Science') over de selectie van personen met fysieke en/of psychische kwetsbaarheid levert heel wat minder studies op in vergelijking met studies van personen die behoren tot de zogenaamde 'Big 3' van diversiteit (leeftijd, geslacht, etniciteit). Hoe kunnen we de relatief beperkte aandacht voor de selectie van personen met een fysieke en/of psychische kwetsbaarheid in een arbeid gerelateerde context² verklaren? Het denkkader dat de personeelsselectie sinds het begin van de vorige eeuw domineert, is het 'klassieke selectieperspectief' waarin het efficiëntie-denken centraal staat (Herriot 1989), zoals hieronder verder toegelicht.

Perspectieven op Selectie

Het klassieke selectieperspectief

Personeelsselectie is één van de belangrijkste HRM activiteiten in organisaties. Toch is de personeelsselectie een vrij jonge discipline. In 1913 publiceerde Hugo Münsterberg zijn boek

¹ Correspondentietests zijn een type praktijktests of experimentele methode waarbij men nagaat of twee of meerdere even gekwalificeerde kandidaten die enkel verschillen van elkaar in een niet-functierelevant kenmerk (bijv. huidskleur of beperking) dezelfde arbeidsmarktkansen krijgen (bijv. in dezelfde mate uitgenodigd worden voor een gesprek). Bij correspondentietests worden sollicitatiebrieven en/of curriculum vitae opgestuurd naar dezelfde werkgever en de 'callback' (mate van uitnodiging) bestudeerd. Er is geen persoonlijk contact en er wordt geen discriminerende vraag gesteld (Deros, Ryan, & Nguyen, 2012; Verhaeghe & Van der Bracht, 2017).

² Er is opvallend meer wetenschappelijke literatuur te vinden over het testen van studenten met een beperking (onderwijscontext) dan het testen van sollicitanten met een beperking (arbeidscontext).

'Psychology and Industrial Efficiency' dat algemeen beschouwd wordt als het eerste wetenschappelijke werk over personeelsselectie (Münsterberg, 1913) en past binnen de 'scientific management' traditie (Taylorisme) waarin een rationele en efficiënte benadering van arbeid voorop staan. Münsterberg's werk heeft een belangrijke stempel gedrukt op de personeelsselectie. Münsterberg promootte niet alleen een wetenschappelijk maar ook een zeer functioneel en praktisch perspectief op personeelsselectie waarbij het de bedoeling was om 'de juiste man op de juiste plek voor de best mogelijke uitkomst', te krijgen. Goede medewerkers zijn inderdaad hun gewicht in goud waard: uit utiliteitsstudies blijken goede medewerkers gemiddeld 40% tot 70% extra 'op te leveren' van het gemiddelde jaarsalaris van hun functie.

Münsterberg zette zo meteen ook de krijtlijnen uit voor een 'evidence-based' benadering van personeelsselectie. Evidence-based management in organisaties verwijst naar *'a decision-making process combining critical thinking with use of the best available scientific evidence and business information.'* (Rousseau & Barends 2011). Een goed selectiebeleid is de facto 'evidence-based' en start met het bepalen van de functie-eisen en selectiecriteria (d.w.z., competenties en eigenschappen) die van belang zijn voor een goed functioneren op de (toekomstige) werkplek. Deze eisen en criteria worden bepaald volgens wetenschappelijk onderbouwde functiebeschrijvingen en competentie-analyses in overleg met belangrijke sleutelfiguren in de organisatie (bijv., directe leidinggevendenden, HR medewerkers). Vervolgens worden de selectiecriteria gemeten door middel van betrouwbare en wetenschappelijk gevalideerde selectie-instrumenten, zoals een psychologische test. Tot slot, wordt onderzocht of de sollicitant voldoet aan de selectiecriteria door na te gaan of zijn/haar testresultaten beantwoorden aan de vooropgestelde normen en verwachtingen. Dit gebeurt eveneens op basis van de best beschikbare wetenschappelijke evidentie en informatie uit het praktijkveld.

Münsterberg's erfenis, het klassieke selectieperspectief, wordt daarnaast gekenmerkt door een aantal assumpties, zoals (a) het vinden van de 'beste kandidaat' voor het 'beste effect', (b) standaardisatie en stabiliteit, en (c) éénrichtingsverkeer (Derous & De Witte 2001). Ten eerste, zijn selectieprocedures net als een hordeloop: sollicitanten moeten succesvol verschillende opgelegde 'hordes' (selectiecriteria) nemen om tot de finish (of baan) te geraken. Traditioneel leggen werkgevers de lat hoog, wat begrijpelijk is vanuit een utilitair denken: werknemers zijn duur en goede medewerkers leiden tot een hogere 'return on investment'. Binnen het klassieke selectieperspectief streeft men immers naar het vinden van de 'beste kandidaat', wat een soort emanatie is van het toenmalige, klassieke darwinistische discours. De hordes (selectiecriteria) kunnen echter zo hoog ingesteld worden dat kandidaten er maar met moeite over geraken. Hoe strenger de selectie (of lager de selectieratio), hoe meer kandidaten

sneuvelen, waaronder ook heel wat goede kandidaten ('positieve missers') die om niet-functierelevante redenen de hordes niet halen (Guion, 2011). Het klassieke selectieperspectief weerspiegelt immers visies op mens en arbeid die het grootste deel van de vorige eeuw kenmerkten. Selectieprocedures werden ontwikkeld door en voor de toenmalige werkende 'middenklasse' die doorgaans 'blank', 'jong', 'valide' en 'mannelijk' was. Bovendien zinderde het darwinistisch gedachtengoed in de eerste helft van de vorige eeuw flink na: de psychotechniek³ vierde hoogtij waarbij het meten van mentale en fysieke fitheid centraal stond. Dit is één mogelijke verklaring voor de beperkte aandacht voor (selectieprocedures voor) kandidaten uit bepaalde kansengroepen, zoals kandidaten met een beperking.

Het klassieke selectieperspectief gaat, ten tweede, uit van stabiliteit en standaardisatie. Selectietests proberen zo goed mogelijk te voorspellen hoe kandidaten het later in de functie zullen doen. Voorspellen kan in principe pas goed als persoons- en omgevingskenmerken relatief stabiel zijn. Factoren die leiden tot gedragsinconsistentie worden binnen dit perspectief beschouwd als meetfouten die te wijten zijn aan de kandidaat (bijv. aandacht niveau; gemoedstoestand), de context (bijv. een lange test of drukke testomgeving) en de interactie tussen beide (bijv. vermoeidheid omwille van een lange test of gebrek aan concentratie door een drukke testomgeving). Het klassieke testmodel beschouwt deze factoren als toevallige meetfouten en houdt hier weinig rekening mee. Vanuit dit denkkader is standaardisatie (d.w.z. alle kandidaten op dezelfde manier behandelen tijdens de selectieprocedure) immers de manier om de voorspellende waarde van selectietests te maximaliseren. De lat wordt hierbij voor iedereen even hoog gelegd: er wordt geen rekening gehouden met de complexiteit van gedrag die bijv. bepaalde beperkingen met zich meebrengen.

Tot slot, gaat het klassieke selectieperspectief uit van éénrichtingsverkeer. "*Although lip service is paid to [candidates'] needs and motives, the emphasis first and foremost is on the fulfilment of the organisation's needs*" (Schein, 1965, p. 25). Zoals Schein reeds in de jaren '60 van de vorige eeuw schreef, bepaalt de organisatie (of werkgever, HR afdeling, werving- en selectieconsulent) vrij eenzijdig wie, wanneer, waar en hoe getest wordt. Het klassieke selectieperspectief houdt weinig rekening met de mogelijkheid, wensen en verwachtingen van de sollicitant, bijvoorbeeld inzake aangepaste testomstandigheden.

Het sociaal procesperspectief

³ Zo beschreef Münsterberg hoe je op een wetenschappelijke manier telefoonoperators, trambestuurders, en zeevaartkapiteins kon selecteren aan de hand van psychotechnische proeven (de 'psychotechniek'). Typische psychotechnische proeven uit die tijd waren reactietijdmetingen, die zo uit de laboratoria van de experimentele psychologen 'weggeplukt' leken.

Vanaf de jaren '80-'90 van de vorige eeuw begint het tij te keren. Academici benaderen personeelsselectie op een andere manier. Veranderingen op de arbeidsmarkt (zoals schaarste en toegenomen diversiteit) leidden tot een alternatief perspectief, het zogenaamde 'sociaal procesperspectief' (Herriot 1989; Iles & Salaman 1995). Dit perspectief gaat uit van drie andere assumpties dan het klassieke selectieperspectief: (a) het vinden van 'een goede' in plaats van 'de best mogelijke' kandidaat, (b) billijkheid/adaptatie, en (c) tweerichtingsverkeer. Deze principes vormen ook de basis voor een meer inclusieve benadering van personeelsselectie.

Ten eerste, pleitten Rynes en Barber (1990) er vanaf het begin van de jaren '90 voor om (ook) open te staan voor kandidaten die misschien niet het best gekwalificeerd zijn maar wel een minimale bagage bezitten om goed te kunnen functioneren in organisaties. Hoe hoog men de hordes instelt, is dan een bredere beleidsvraag die verder reikt dan een beslissing van de selectiespecialist maar ook weerspiegelt hoe organisaties omgaan met diversiteit en validiteit. Het is bijvoorbeeld aan organisaties om te bepalen of functierelevante competenties perfect beheerst moeten worden bij intrede of verder ontwikkeld kunnen worden op de werkplek. In sommige omstandigheden is het raadzaam om erg selectief te zijn en moeten kandidaten een competentie (zoals een taal of andere vaardigheid) perfect beheersen vanuit functierelevante overwegingen (bijv. wanneer veiligheid gegarandeerd moet worden). In andere omstandigheden is een minder perfecte beheersing van competenties verdedigbaar als die competenties niet cruciaal zijn. De functie blijft de maatstaf die het beleid rechtvaardigt.

Ten tweede, houdt het sociaal procesperspectief rekening met billijkheid en aanpasbaarheid van procedures (Derous & De Witte 2001). Vanuit een meer inclusief denken kan men zich afvragen of er omstandigheden zijn waarin een meer 'billijke' behandeling ('equity') te verkiezen is boven een meer 'gelijke' behandeling ('equality'). In het klassieke selectiedenken staat het aanpassen van de kandidaat aan de test (en functiecriteria) centraal. Binnen een meer inclusieve benadering van personeelsselectie ligt de nadruk eerder op een redelijke aanpassing van de test (en de functiecriteria) aan de kandidaat. Aanpassingen kunnen gebeuren op verschillende niveaus, zoals de beoordelaar, het instrument (testitems; instructies; antwoordmodaliteiten) en zelfs de functie (functie-eisen; functie-omgeving). Schriftelijke tests die een bepaalde leesvaardigheid vereisen, kunnen bijvoorbeeld een struikelblok zijn voor kandidaten met een andere taalachtergrond dan de taal waarin de test is opgesteld of kunnen moeilijker te begrijpen en in te vullen zijn voor kandidaten met dyslexie. Deze tests zouden ook op een andere manier aangeboden kunnen worden die rekening houdt met de beperkingen van de kandidaat en een goede functie-uitoefening niet in de weg staan (bijv. door te werken aan de begrijpbaarheid van de testitems). Een belangrijke vraag bij een meer inclusieve benadering

van selectie is hoe testprocedures en –bij uitbreiding- functies aangepast kunnen worden aan de specifieke noden van kandidaten.

Tot slot, gaat het sociaal procesperspectief uit van tweerichtingsverkeer. Organisaties houden rekening met de mening, perceptie en reactie van sollicitanten op selectiepraktijken, zoals gepercipieerde rechtvaardigheid (McCarthy et al., 2017) en discriminatie (Anderson, 2011). Om onterechte scoreverschillen in tests te reduceren, kunnen kandidaten testtrainingen krijgen of zelfs hertest worden (Schleicher, Van Iddekinge, Morgeson, & Campion, 2010). Ook kunnen potentiële kandidaten betrokken worden bij de ontwikkeling of het aanpassen van bestaande tests, om bijvoorbeeld onduidelijkheden in testinstructies of testitems weg te werken (Oostrom & Born 2014).

Maar testaccommodatie roept ook vragen op. Zijn de aanpassingen aan tests (testinhoud, testinstructies, antwoordmodaliteiten) verzoenbaar met de objectiviteit en validiteit van testprocedures? In de volgende paragraaf bestuderen we de meest voorkomende selectietests en de uitdagingen die redelijke aanpassingen stellen aan twee belangrijke kwaliteitscriteria in het bijzonder, nl. de validiteit (voorspellende waarde) en fairness (eerlijkheid) van selectietests.

Psychologische tests, Validiteit, Fairness en Testaanpassingen

Soorten tests

Er wordt doorgaans een onderscheid gemaakt naar twee grote groepen van selectie-instrumenten, nl. psychologische tests en de tests die gebruikt worden ter voorbereiding op het feitelijke psychologische onderzoek. *Psychologische tests* zijn een verzameling van opdrachten of vragen die specifiek gedrag uitlokken onder gestandaardiseerde omstandigheden en een voldoende onderscheidend vermogen hebben om een objectieve vergelijking tussen personen mogelijk te maken (Roe, 1983). Deze selectie-instrumenten of tests worden ingedeeld op basis van het selectieconstruct dat ze meten, zoals cognitieve vaardigheden of intelligentie (nl. intelligentietests, ook cognitieve vaardigheidstests of redeneertests genoemd), persoonlijkheid en waarden (nl. persoonlijkheidsvragenlijsten; waarden- en interessevragenlijsten), en gedrag (nl. objectieve en subjectieve gedragsproeven). Het selectie-interview, daarentegen, is een meer ‘holistisch’ of ‘heterogeen’ meetinstrument dan de andere psychologische tests omdat verschillende psychologische constructen tegelijkertijd gemeten worden, zoals iemands’ persoonlijkheid, interesses en competenties. Daarnaast is er nog een groep selectie-instrumenten die gebruikt wordt als voorbereiding op het psychologische onderzoek, zoals het Curriculum Vitae (CV’s) en referentiechecks. Deze tests zijn in strikte zin géén psychologische tests maar worden soms (impliciet/onbewust) gebruikt als psychologische test. Zo leiden

recruiters op basis van CV-gegevens aspecten van iemands persoonlijkheid af, ook al lenen CV's zich in principe enkel voor een eerste screening op basis van objectief vaststelbare, biografische gegevens zoals als iemands' vooropleiding, type en hoeveelheid werkervaring, enz. (zie Apers & Derous 2017; Derous & Decoster 2017).

Kwaliteitseisen

Psychologische tests moeten aan een groot aantal kwaliteitseisen voldoen wat betreft het *testmateriaal*, de *testhandleiding*, de *normering* (om testresultaten van kandidaten te vergelijken met referentiegroepen), de *betrouwbaarheid* (of testresultaten stabiel zijn over testmomenten en beoordelaars heen), de *theoretische inbedding* (welke psychologische theorie aan de grondslag ligt van het construct dat gemeten wordt in de psychologische test), de *inhoudsvaliditeit* (of de testitems een representatieve afspiegeling zijn van het domein dat bevraagd wordt) en de *constructvaliditeit* (of hoe goed de test het psychologische construct meet waarvoor de test ontwikkeld is en geen andere constructen meet) (zie Evers et al. 2010). De *criteriumgerelateerde validiteit* (of voorspellende waarde) gaat over de mate waarin de testresultaten samenhangen met het criterium dat men wil voorspellen (zoals toekomstige functieprestaties) en deze samenhang wordt uitgedrukt met een validiteitscoëfficiënt⁴. Standaardisatie of gelijke behandeling is een belangrijke voorwaarde om de voorspellende validiteit van tests te garanderen (zie 'Het klassieke selectieperspectief'). *Test fairness*, tot slot, is een complex fenomeen. Diverse aspecten van een psychologische test bepalen samen de fairness van een test (American Educational Research Association, American Psychological Association, 2014). Test fairness is dus geen geïsoleerd kenmerk van een test. Toch komt het onderzoek naar de testfairness grosso modo neer op het bestuderen van de construct- en de criteriumgerelateerde validiteit van de test. Volgens de Standards for educational and Psychological Testing (2014) is fairness "a fundamental validity issue that requires attention throughout all stages of test development and use" (p. 49). In de praktijk beschouwt men een test als fair (of eerlijk) als de interne structuur van de test (of het beoogde construct) niet verschilt voor verschillende doelgroepen, als er geen vraagpartijdigheid is (waarbij kandidaten uit verschillende groepen anders reageren op een vraag of testitem) en als er geen aanzienlijke (onterechte) verschillen in testcores en selectieratio's⁵ zijn tussen kandidaten uit verschillende groepen (bijv. mannen vs. vrouwen, etnisch-culturele minderheden vs. meerderheden, oudere

⁴ Een validiteitscoëfficiënt r ligt tussen 0 (geen verband) en 1 (perfect verband) waarbij de hoogste waarde van validiteitscoëfficiënten van psychologische tests doorgaans rond de .60-.70 liggen.

⁵ De selectieratio is het percentage van kandidaten dat slaagt op een test.

vs. jongere kandidaten, kandidaten met en zonder beperkingen). Naast ‘objectieve’ test fairness, kunnen kandidaten ook het idee of gevoel hebben dat een test niet fair is. Objectieve en subjectieve test fairness kunnen maar moeten derhalve niet samenvallen (Derous & Ryan 2014). Zo kan een psychologische test objectief gezien fair zijn maar kunnen kandidaten zich niet fair behandeld voelen en omgekeerd.

Twee kwaliteitseisen die bijzonder relevant zijn bij inclusief selecteren, zijn de validiteit en fairness van tests. We bespreken daarom in de volgende paragraaf wat algemeen geweten is over de voorspellende waarde en eerlijkheid van de belangrijkste psychologische tests. Wat fairness betreft, spitsen we ons toe op objectieve eerlijkheid zoals gemeten via differentiële testcores en selectieratio's.

Validiteit en fairness van tests

Algemene intelligentietests en gedragsproeven hebben gemiddeld genomen de hoogste voorspellende waarde van alle psychologische tests. *Algemene intelligentietests* meten iemands vermogen om complexe informatie te verwerken en toe te passen op –doorgaans- nieuwe situaties. Deze tests bestaan meestal uit een aantal subtests die specifieke vermogens meten (zoals numeriek redeneervermogen, verbaal redeneervermogen of perceptuele snelheid in het verwerken van informatie). *Gedragsproeven*, daarentegen, zijn situationele tests: ze confronteren sollicitanten met concrete situaties (of taken) die een weerspiegeling zijn van de toekomstige functie waarvoor men solliciteert. Een voorbeeld van een gedragsproef is een postbakoefening waarbij de kandidaat op basis van documenten die in zijn/haar postbak terecht vragen beantwoordt en een agenda plant. Een situationele beoordelingstest (SJT) is een contextueel selectie instrument waarbij kandidaten een reeks korte, functierelevante situaties die peilen naar functierelevant gedrag lezen (op papier) of bekijken (op video) en elke situatie moeten beoordelen aan de hand van een reeks voor gestructureerde antwoordalternatieven of zelf een antwoord bedenken. Meta-analyses tonen een gemiddelde voorspellende waarde van $r = .48$ tot $r = .65$ voor algemene intelligentietests, van $r = .33$ tot $r = .54$ voor gedragsproeven en rond de $r = .26$ voor situationele beoordelingstests (Schmidt & Hunter 1998; Schmidt & Hunter 2004). Naarmate een functie complexer is, voorspellen algemene intelligentietests toekomstige werkprestaties zelfs beter. Maar in tegenstelling tot gedragsproeven, leiden algemene intelligentietests ook tot aanzienlijk grotere subgroepverschillen in het nadeel van kandidaten uit bepaalde groepen zoals etnische minderheidsgroepen en oudere medewerkers (er is gemiddeld genomen geen significant verschil naar geslacht). Sommige tests, zoals algemene intelligentietests hebben dus een zeer hoog voorspellende waarde maar leiden ook tot

zeer grote subgroepverschillen. Omdat het realiseren van validiteit op gespannen voet staat met het realiseren van diversiteit, wordt dit het ‘diversiteit-validiteit’ dilemma genoemd. Gedragstoeetsen, daarentegen, vertonen erg weinig subgroepverschillen naar geslacht, leeftijd en etniciteit en zijn gemiddeld genomen zeer valide en doorgaans ook erg billijk (Hough, Oswald, & Ployhart 2001).

Een andere, populaire psychologische test is de persoonlijkheidsvragenlijst. *Persoonlijkheidsvragenlijsten* meten de persoonlijkheid van kandidaten aan de hand van vijf grote persoonlijkheidsdimensies of –trekken, ook het vijf-factoren of ‘Big-five’ model⁶ genoemd: Conscientieusheid (of gewetensvolheid), Vriendelijkheid, Openheid voor ervaringen, Emotionele stabiliteit en Extraversie. Deze vijf factoren bestaan elk uit verschillende deelfactoren of facetten. De voorspellende waarde van persoonlijkheidsvragenlijsten varieert verder per persoonlijkheidsfactor (en zelfs facet) en per criterium. Voor algemene functieprestaties is de voorspellende waarde het grootst voor Conscientieusheid, ($r = .22$), gevolgd door Emotionele stabiliteit ($r = .12$), Extraversie ($r = .09$), Vriendelijkheid ($r = .08$), en Openheid voor ervaringen ($r = .04$) (Schmidt, Shaffer, & Oh 2008; Schmidt & Hunter 1998). In tegenstelling tot algemene intelligentietests en gedragstoeetsen, is de voorspellende waarde van persoonlijkheidsvragenlijsten aanzienlijk kleiner maar zijn er amper scoreverschillen naar etnische origine. Vrouwen scoren op sommige persoonlijkheidsfacetten gemiddeld genomen iets hoger dan mannen (zoals bij Vriendelijkheid) en oudere personen scoren in vergelijking met jongere personen op sommige factoren gemiddeld ook wat hoger (zoals Conscientieusheid) en op andere facetten gemiddeld genomen dan weer iets lager (zoals bij Achievement).

Het *selectie interview* is, tot slot, een gesprek waarbij één of meerdere interviewers aan kandidaten vragen stellen en antwoorden evalueren om de functiegeschiktheid van de kandidaten in te schatten (Levashina, Hartwell, Morgeson, & Campion, 2014). Naast het toetsen van geschiktheid, worden interviews ook gebruikt voor andere doeleinden, zoals het uitwisselen van informatie en het leggen van persoonlijk contact. Er bestaat grote consensus onder academici over het belang en de waarde van gestructureerde interviews (dezelfde functie-relevante vragen, in dezelfde volgorde, met dezelfde interviewers, beoordelingsschalen). Vrijwel alle gepubliceerde meta-analyses die een ‘peer-review’ ondergingen (Schmidt & Hunter 1998), tonen dat de validiteit van gestructureerde interviews, zoals het gedragsgerichte

⁶ Er is recent meer en meer wetenschappelijke evidentie voor een zes-factoren model van persoonlijkheid waarin ‘Integriteit-Humiliteit’ als zesde persoonlijkheidsfactor is opgenomen (zie de Vries, Ashton, & Lee 2009).

of situationele interview ($r = .51$) groter is dan de validiteit van ongestructureerde interviews ($r = .38$). Structuur zorgt ervoor dat vragen functierelevant zijn en dat stoorzenders zoveel mogelijk onder controle gehouden worden. De kans op bias is immers het grootst wanneer interviews ongestructureerd verlopen en kandidaten bovendien één of ander (zichtbaar) stigma hebben, zoals een andere huidskleur, oudere leeftijd, obesitas of fysieke beperking (Derous, Buijsrogge, Roulin, & Duyck, 2016).

Validiteit, fairness en testaanpassingen

Een erg belangrijke vraag is hoe we diversiteit (op basis van een faire testafname voor verschillende doelgroepen) kunnen realiseren zonder in te boeten aan de validiteit van psychologische tests. Academics bestuderen al geruime tijd strategieën om uit het diversiteit-validiteit dilemma te geraken. De meeste studies richtten zich op etniciteit. Zo zijn er, ten eerste, *op de persoon gerichte strategieën*, zoals het hertesten van sollicitanten (Schleicher et al., 2010), het voorzien van testvoorbereiding en –training (Sackett, Schmitt, Ellingson, & Kabin, 2001), het verhogen van de testmotivatie van sollicitanten en het controleren voor demografische kenmerken (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997). De effectiviteit van deze strategieën is eerder beperkt, tenminste wanneer we kijken naar strategieën om etnische subgroepverschillen te reduceren (d.w.z. deze strategieën leiden tot kleine reducties in etnische subgroepverschillen). Ten tweede, zijn er ook *op de beoordelaar* (HR manager; selectiespecialist; recruiter) *gerichte strategieën*, zoals het aanbieden van diversiteitstrainingen. Beoordelaars kunnen een opleiding volgen om beoordelingsfouten of vooroordelen te vermijden. Doorgaans zijn deze effecten eveneens klein en zakken ze vrij snel weg na de interventie (Derous & Ryan 2014). Tot slot, zijn er ook *op het selectie-instrument* gerichte strategieën om (onterechte) subgroepverschillen te reduceren. Hier wordt het meeste onderzoek naar gedaan. Volgens het klassieke selectieperspectief zijn standaardisatie van testafname en testinhoud erg belangrijk om tot betrouwbare testresultaten te komen zodat men kandidaten goed kan vergelijken en job prestaties optimaal kan voorspellen. Maar niet alle kandidaten zijn in staat om even goed te presteren onder gestandaardiseerde testcondities (qua testafname en -inhoud). Om etnische subgroepverschillen te reduceren, minimaliseert men bijvoorbeeld de nodige taalkennis in vragen en vraaginstructies en bestudeert men alternatieve antwoordmodaliteiten (De Soete, Lievens, Oostrom, & Westerveld, 2013), verwijdert men tijdslimieten (Sackett et al., 2001) en gebruikt men bijvoorbeeld selectie-instrumenten die een grotere verscheidenheid aan kennis en vaardigheden meten dan cognitieve vaardigheden (De Corte, Lievens, & Sackett 2007).

Tot nu toe richtten de meeste wetenschappelijke studies zich op strategieën om etnische subgroep verschillen te reduceren. Echter, redelijke testaanpassingen kunnen ook nodig zijn voor kandidaten uit andere doelgroepen, zoals mensen met een fysieke beperking of psychische kwetsbaarheid, bijvoorbeeld. Door redelijke aanpassingen aan tests door te voeren, wil men de testprestaties van een kandidaat corrigeren voor diverse persoonsgebonden kenmerken (zoals beperkingen) die géén voorspellende waarde hebben voor het toekomstige functioneren van die kandidaat. Maar met testaanpassingen moet men voorzichtig omspringen. Testaanpassingen voor specifieke doelgroepen impliceren dat de test niet meer gestandaardiseerd is, tenzij deze aanpassingen consistent en consequent doorgevoerd worden voor iedereen. Door een gebrek aan standaardisatie wordt het moeilijker om kandidaten op een objectieve manier met elkaar te vergelijken en is in principe de testvaliditeit niet meer gegarandeerd (Roe, 1983).

Zijn alle testaanpassingen nefast? In de literatuur wordt een onderscheid gemaakt tussen testaccommodatie en testmodificatie (Stone, Laitusis, & Cook 2015). Testaccommodatie wijzigt de manier waarop de test wordt afgenomen zonder dat de (construct)validiteit van de test in het gedrag komt. Voorbeelden van testaccommodatie zijn: het wijzigen van de presentatie- en antwoordmodaliteiten (bijv. het voorlezen van de vragen of het aanbieden van een open antwoordmogelijkheid i.p.v. gesloten antwoordmogelijkheid; De Soete et al. 2013). Of testaccommodatie nefast is voor de validiteit is dus een empirische vraag die afhankelijk is van het type accommodatie, het type construct dat gemeten wordt en de concrete doelgroep waarvoor de testaccommodatie bedoeld is. Maar ook al doet een testaanpassing geen afbreuk aan de constructvaliditeit, dan nog kunnen testaanpassingen problematisch zijn wanneer ze een onterecht voordeel bieden aan kandidaten uit bepaalde doelgroepen (bijv. aan kandidaten vanuit een etnisch-culturele minderheidsgroep, met een zekere leeftijd of met een bepaalde beperking) in vergelijking met de kandidaten uit andere doelgroepen (zoals kandidaten uit een etnisch-culturele meerderheidsgroep, die jong zijn of zonder beperking; Camilli 2013). Daarmee samenhangend, mogen testaccommodaties niet compenseren voor competenties of vaardigheden die de kandidaat niet bezit of verworven heeft. Tot slot, kunnen testaanpassingen problematisch zijn als ze geen goede afspiegeling zijn van de latere functie-eisen. Stel dat de aanpassingen wél aan de test maar níet op de werkplek doorgevoerd worden, dan is de test (en testprestatie) niet (meer) functierelevant, kan de testaanpassing leiden tot een overschatting van het potentieel van de kandidaat en dus een lagere voorspellende waarde. Bij testmodificatie, daarentegen, zal het construct dat gemeten wordt door de testaanpassingen zélf veranderen waardoor ook de voorspellende waarde en eerlijkheid van de test in het gedrang komen. Men meet dus iets anders dan dat wat de test oorspronkelijk bedoelde te meten of men meet iets dat

niet (meer) relevant is voor de latere functie uitoefening en dit heeft een averechts effect op de algemene kwaliteit van de testafname.

Wanneer zijn testaanpassingen dan wel aanvaardbaar? Dat is een pittige vraag. In de wetenschappelijke literatuur bestaan hierover twee visies. Een eerste (strengere) visie stelt dat door de testaanpassingen de testcores van kandidaten die geen testaanpassing nodig hebben niet mogen verbeteren. Enkel de testcores van kandidaten die de aanpassing nodig hebben, mogen verbeteren. Dit wordt de ‘interactiehypothese’ of ‘maximal potential hypothesis’ genoemd (zie Gregg & Nelson 2012). Een tweede, minder strenge visie stelt dat de verbetering in testcores van kandidaten die de aanpassing nodig hebben significant groter moet zijn dan de verbetering in de testcores van de kandidaten die de testaanpassing niet nodig hebben. Dit wordt ook de ‘differential boost’ hypothese genoemd (Fuchs & Fuchs 1999). Kortom: men accepteert dat de testaanpassing een voordeel oplevert voor iedere kandidaat maar dit voordeel moet substantieel groter zijn voor de kandidaat die de testaanpassing effectief nodig heeft. Maar hoeveel groter dit voordeel dan moet zijn, is niet zo duidelijk (Cawthon, Ho, Patel, Potvin, & Trundt, 2009). Een bijkomende moeilijkheid bij het bestuderen van testaanpassingen, bijvoorbeeld voor mensen met een beperking, is dat er erg veel variabiliteit is binnen dezelfde categorie van beperkingen. Zo kan men in verschillende gradaties een visuele beperking hebben of leiden aan autisme. Hoe specifiek de doorgevoerde testaanpassingen zijn (d.w.z. hoe meer op maat van specifieke beperkingen en personen), hoe moeilijker het is om voldoende kandidaten te vinden met die beperking (d.w.z. hoe kleiner de groepen kandidaten met die specifieke beperking), hoe lastiger het is om verbanden te vinden (statistisch) en hoe moeilijker het wordt om bevindingen te generaliseren. Kortom, wetenschappelijk onderzoek naar de effectiviteit van testaanpassingen is nodig maar niet zo eenvoudig uit te voeren. In de volgende paragraaf illustreren we met een voorbeeld hoe de effectiviteit van één type redelijke aanpassing aan drie tests (nl. het toekennen van ‘extra tijd’) werd onderzocht.

Illustratie

In het project ‘redelijke aanpassingen’ van Selor - FOD BOSA⁷ onderzochten we testaanpassingen voor kandidaten met specifieke functiebeperkingen, zoals visuele beperkingen. De algemene doelstelling was om een wetenschappelijk onderbouwde beslissingstool te ontwikkelen voor het toekennen van redelijke aanpassingen per type beperking die zowel met de typologieën, de redelijke aanpassing, als de validiteit van de tests rekening hield. Er werd zoveel mogelijk ‘evidence-based’ gewerkt. Zo bestond de werkgroep

⁷ Selor | FOD BOSA is het selectiebureau voor de federale overheid in België.

‘visuele beperkingen’ naast de werkgroep ‘autisme’ uit wetenschappelijke experts, medewerkers van de onderzoeks- en diversiteitsteams van Selor en ervaringsdeskundigen (mensen met beperkingen). In totaal werden drie subwerkgroepen georganiseerd die respectievelijk (a) kennis maakten met alle tests uit de generieke screeningsmodule van Selor, (b) onderzoek deden naar relevante typologieën rond de beperking die het onderwerp uitmaakt van de studie (op basis van ervaring en literatuur) en (c) zich toelegden op een empirische studie naar de effectiviteit van redelijke testaanpassingen.

Omdat op basis van de literatuurstudie en het kwalitatieve onderzoeksgedeelte een aantal onduidelijkheden bleven bestaan over de invloed van iemands’ leessnelheid op testprestaties en de link met het construct algemene cognitieve vaardigheid, werden een experimentele studie (met pre- en posttest enquêtes) opgestart om testaanpassing bij visuele stoornissen verder te onderzoeken. Voor het bestuderen van de effecten van testaanpassingen (zoals ‘extra tijd’, maar ook het gebruik van brailleleesregel, voorleessoftware en zoomtekst ⁸) op de leessnelheid en de testprestaties van deelnemers met een visuele beperking, werd een gemengd factorieel design opgezet waarbij een groep deelnemers met een visuele beperking en een groep deelnemers zonder visuele beperking elk drie verschillende psychologische tests (nl. een test voor abstract redeneervermogen, een postbakoefening en een SJT) éénmaal aflegden met ‘extra tijd’ en éénmaal aflegden zonder ‘extra tijd’. Deelnemers (N=187) werden gerekruteerd via belangenorganisaties (Brailleliga, La Lumière, Spermalì) en via het eigen netwerk. Om leereffecten zoveel mogelijk uit te schakelen, werden de tests met en zonder aanpassingen op twee verschillende momenten afgenomen (met een ontwikkelingsbuffer van 6 maanden tussen het eerste en tweede afnamemoment) en werden de tests (met extra tijd vs. zonder extra tijd) gecontrabalanceerd aangeboden waarbij de helft van de deelnemers eerst de tests met extra tijd en de andere helft van de deelnemers eerst de tests zonder extra tijd invulden. Op die manier kon worden nagegaan of de testscores verbeterden door extra tijd toe te kennen en of deze verbetering in de groep deelnemers met visuele beperking sterker was dan in de groep deelnemers zonder visuele beperking.

Leessnelheid hing positief samen met de prestaties op de abstracte redeneertest maar niet met testprestaties op de postbakoefening en de SJT. Verder medieerde (of verklaarde) leessnelheid het effect van de andere redelijke testaanpassingen (zoals brailleleesregel,

⁸ In een pre- en posttest vragenlijst werden controlevariabelen opgenomen, die polsten naar cognitieve vermoeidheid, voorgaande testervaringen, beperking (soort, aangeboren of niet, kunnen werken met bepaalde software voor slechtzienden, etc.) en ook demografische factoren (zoals geslacht en leeftijd). Bij de pre-test vragenlijst kreeg men een tekst voorgelegd om gemiddelde leessnelheid te bepalen (zie Selor, 2017).

voorleessoftware, en zoomtekst) op de testprestatie op de abstract redeneertest indien geen extra tijd werd toegekend. Wie gebruik maakte van een brailleleesregel (of voorleessoftware of zoomtekst) las trager en scoorde lager op de abstracte redeneerproef (Selor 2017). Testaanpassingen die noodzakelijk zijn voor deelnemers met een visuele beperking verlaagden dus de leessnelheid en testprestaties, waardoor het gerechtvaardigd leek om voor deze groep extra tijd te voorzien.

Om verder na te gaan of ‘extra tijd’ toekennen aan kandidaten met een visuele beperking (die specifieke hulpmiddelen gebruik) wel degelijk gerechtvaardigd was, werd de ‘differential boost’ hypothese onderzocht. De ‘differential boost hypothese’ werd bevestigd voor de abstracte redeneertest maar niet voor de postbakoefening en ook niet voor de SJT. Met andere woorden, voor de postbakoefening en de SJT kon men –op basis van deze studieresultaten- niet garanderen dat de testaanpassing (het toekennen van extra tijd aan kandidaten met een visuele beperking) leidde tot een differentieel en betekenisvol voordeel voor kandidaten met een visuele beperking t.a.v. kandidaten zonder visuele beperking. Omdat snelheid voornamelijk een rol speelt bij de redeneertest en niet bij de postbakoefening en de SJT, zou men op basis van deze onderzoeksbevindingen kunnen beslissen om de tijdslimieten volledig weg te laten bij die laatste twee tests en extra tijd te geven aan mensen met een visuele beperking bij de redeneertest.

Maar om helemaal zeker te zijn dat deze testaanpassing geoorloofd is, zou men nog verder moeten onderzoeken of het toekennen van extra tijd aan mensen met een visuele beperking voor het afleggen van de redeneerproef het onderliggende construct ‘abstract redeneervermogen’ niet wijzigt. Met andere woorden: de toegekende extra tijd mag niet interageren met abstract redeneervermogen als construct. Onderzoek toont bijvoorbeeld dat abstract redeneervermogen samenhangt met perceptuele snelheid (Kranzler & Jensen 1989). Leessnelheid is geen perceptuele snelheid maar perceptuele snelheid werd niet gemeten.

Voor de postbakoefening en de SJT werden er verder geen significant verschillen gevonden op basis extra toegekende tijd en zou men de tijdslimiet kunnen loslaten en alle kandidaten dan verder gewoon de tijd geven die ze nodig achten om die tests op te lossen. De groep deelnemers was uiteraard klein, waardoor de kans op betekenisvolle verbanden (statistisch gezien) niet groot is. Indien mogelijk, is vervolgonderzoek op grotere groepen daarom wenselijk. Maar los daarvan moeten selectieprocedures niet alleen effectief maar ook efficiënt zijn, zeker wanneer een groot aantal kandidaten (zoals bij overheidsselecties) beoordeeld moeten worden. Omwille van deze inhoudelijke en ook praktisch/organisatorische redenen werd de tijdslimiet uiteindelijk wel behouden bij de afname van de postbakoefening en de SJT.

Selectie voor Diversiteit

Testaanpassingen moeten omzichtig gebeuren. Functierelevante criteria moeten de ultieme toetssteen blijven om onterechte discriminatie te vermijden. Heel wat studies, zoals tonen aan dat het verzoenen van diversiteit (maximaliseren van gelijke kansen) en validiteit (maximaliseren van doelmatigheid) niet gemakkelijk is. Er bestaat tot op vandaag geen perfecte, objectieve oplossing voor het diversiteit-validiteit dilemma, tenminste wanneer men vertrekt van een test die volgens bepaalde principes (zoals het klassieke selectieperspectief) en voor bepaalde doelgroepen (zoals personen zonder beperkingen) ontwikkeld werd.

Een meer inclusief selectiebeleid zou ook op een andere manier tot stand kunnen komen. Bij testaccommodatie vertrekt men van een bestaande test die oorspronkelijk niet aangepast werd aan de noden van kandidaten uit specifieke doelgroepen en die men nadien wil aanpassen aan de noden van deze kandidaten, zoals geïllustreerd. In plaats van bestaande tests aan te passen zou men volgens het principe van ‘Universal Design’ (Mace, Hardie, & Place 1996) nieuwe tests of selectieprocedures van meet-af aan zo kunnen ontwerpen dat ze optimaal geschikt zijn voor zoveel mogelijk verschillende kandidaten en doelgroepen. Zo zou men bijvoorbeeld een volledig nieuwe abstracte redeneertest kunnen ontwikkelen in 3D die zowel bij personen met als zonder visuele beperking op een even betrouwbare en valide manier het abstract redeneervermogen meet, zonder dat er nood is aan het toekennen van extra tijd voor bepaalde kandidaten (Selor, 2017).

Het ‘Universal Design’ principe werd voor het eerste uitgewerkt in de architectuur en vertrekt vanuit het idee dat leef- en werkomgevingen, producten en diensten zo ingericht kunnen worden dat ze bruikbaar en valide zijn voor iedereen. Dit principe is sindsdien al toegepast op verschillende domeinen, waaronder ook het inrichten van leeromgevingen en evaluatiemomenten in een onderwijscontext (Thompson, Johnstone, & Thurlow 2002). Het ‘Universal Design’ principe zou ook toegepast kunnen worden op selectie- en assessment procedures in een werkcontext. ‘Universal Design’ voor selectie zou dan betekenen dat tests en testprocedures niet alleen gekenmerkt worden door een grotere toegankelijkheid voor een erg diverse groep aan kandidaten, maar ook door het meten van goed afgelijnde constructen met vragen die goed verstaanbaar zijn, geen itembias hebben, en waarbij er eenvoudige, duidelijke en intuïtief-begrijpbare instructies en procedures zijn, die minimale testaccommodatie vereisen. Selectietests en -procedures die ontwikkeld zijn volgens dit principe verkleinen de nood aan testaccommodatie (en -modificatie) door allerlei mogelijke barrières voor zoveel mogelijk kandidaten zoveel mogelijk weg te werken. Zelfs als er specifieke accommodatie nodig is (bijv.

gebruik van een brailleleesregel) zouden selectietests zo ingericht kunnen worden dat ze voor een zo breed mogelijk publiek toegankelijk blijven (bijv. door onnodige grafieken weg te laten).

Samengevat, door de toegenomen diversiteit op de arbeidsmarkt is er nood aan een meer inclusief selectiebeleid. Testaccommodatie kan een eerste, goede stap zijn voor het realiseren van inclusieve selectieprocedures maar het tegelijkertijd realiseren van test validiteit en test fairness blijft een grote uitdaging. Inclusief selecteren is meer dan testaccommodatie. Renoveren is goed maar soms is het beter om van nul-af-aan iets nieuws op te bouwen. Volgens het principe van ‘Universal Design’ zou men ook tests van nul-af-aan kunnen uitdenken en ontwikkelen die toegankelijk, valide en fair zijn voor zoveel mogelijk verschillende kandidaten en op die manier zowel principes van het klassieke selectieperspectief als sociale procesperspectief met elkaar verzoent.

Bibliografie

- American Educational Research Association, American Psychological Association, & N. C. on M. in E. (2014). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
- Anderson, N. (2011). Perceived job discrimination: Toward a model of applicant propensity to case initiation in selection. *International Journal of Selection and Assessment*, 19(3), 229–244. doi:10.1111/j.1468-2389.2011.00551.x
- Apers, C., & Derous, E. (2017). Are they accurate? Recruiters’ personality judgments in paper versus video resumes. *Computers in Human Behavior*, 73, 9–19. doi:10.1016/j.chb.2017.02.063
- Camilli, G. (2013). Ongoing issues in test fairness. *Educational Research and Evaluation*, 19(2-3), 104–120. doi:10.1080/13803611.2013.767602
- Cawthon, S. W., Ho, E., Patel, P. G., Potvin, D. C., & Trundt, K. M. (2009). Multiple constructs and the effects of accommodations on standardized test scores for students with disabilities. *Practical Assessment, Research & Evaluation*, 14(18). Retrieved from <http://pareonline.net/getvn.asp?v=14&n=18>
- Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, 82(2), 300–310. doi:10.1037/0021-9010.82.2.300
- De Corte, W., Lievens, F., & Sackett, P. R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology*,

- 92(5), 1380–1393. doi:10.1037/0021-9010.92.5.1380
- De Soete, B., Lievens, F., Oostrom, J., & Westerveld, L. (2013). Alternative predictors for dealing with the diversity-validity dilemma in personnel selection: The constructed response multimedia test. *International Journal of Selection and Assessment*, 21(3), 239–250. doi:10.1111/ijsa.12034
- de Vries, R. E., Ashton, M. C., & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en HEXACO Persoonlijkheidsvragenlijst. *Gedrag & Organisatie*, 22, 232–274.
- Derous, E., Buijsrogge, A., Roulin, N., & Duyck, W. (2016). Why your stigma isn't hired: A dual-process framework of interview bias. *Human Resource Management Review*, 26(2), 90–111. doi:10.1016/j.hrmr.2015.09.006
- Derous, E., & De Witte, K. (2001). Looking at selection from a social process perspective: Towards a social process model on personnel selection. *European Journal of Work and Organizational Psychology*, 10(3), 319–342. doi:10.1080/13594320143000708
- Derous, E., & Decoster, J. (2017). Implicit age cues in resumes: Subtle effects on hiring discrimination. *Frontiers in Psychology*, 8. doi:10.3389/fpsyg.2017.01321
- Derous, E., & Ryan, A. M. (2014). By any other name: Discrimination in resume screening. In U. Klehe & E. a J. van Hooft (Eds.), *The Oxford Handbook of job loss and job search*. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199764921.013.017
- Derous, E., Ryan, A. M., & Nguyen, H. H. D. (2012). Multiple categorization in resume screening: Examining effects on hiring discrimination against Arab applicants in field and lab settings. *Journal of Organizational Behavior*, 33(4), 544–570.
- Evers, A., Lucassen, W., Meijer, R., & Sijsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN.
- Fraser, R. T., Johnson, K., Hebert, J., Ajzen, I., Copeland, J., Brown, P., & Chan, F. (2010). Understanding Employers' hiring intentions in relation to qualified workers with disabilities: Preliminary findings. *Journal of Occupational Rehabilitation*, 20(4), 420–426. doi:10.1007/s10926-009-9220-1
- Fuchs, L. S., & Fuchs, D. (1999). Fair and unfair testing accommodations. *School Administrator*, 56, 24–29.
- Gregg, N., & Nelson, J. M. (2012). Meta-analysis on the effectiveness of extra time as a test accommodation for transitioning adolescents with learning disabilities. *Journal of Learning Disabilities*, 45(2), 128–138. doi:10.1177/0022219409355484
- Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions*. New

- York: Routledge.
- Herriot, P. (1989). Selection as a social process. In M. Smith & I. T. Robertson (Eds.), *Advances in Selection and Assessment* (pp. 171–187). Chichester, UK: Wiley.
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, 9(1&2), 152–194. doi:10.1111/1468-2389.00171
- Iles, P. A., & Salaman, G. (1995). Recruitment, selection, and assessment. In J. Storey (Ed.), *Human Resource Management: A critical text* (pp. 203–233). London: Routledge.
- Kranzler, J. H., & Jensen, A. R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence*, 13(4), 329–347. doi:10.1016/S0160-2896(89)80006-6
- Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, 67(1), 241–293. doi:10.1111/peps.12052
- Mace, R. L., Hardie, G. J., & Place, J. P. (1996). *Accessible environments: Toward universal design*. North Carolina: The Center for Universal Design.
- McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing “So what?,” “What’s new?,” and “Where to next?” *Journal of Management*, 43(6), 1693–1725. doi:10.1177/0149206316681846
- Münsterberg, H. (1913). *Psychology and Industrial Efficiency*. Boston/New York: Houghton Mifflin.
- Neumark, D., Burn, I., & Button, P. (2016). Experimental age discrimination evidence and the Heckman critique. *American Economic Review*, 106(5), 303–308. doi:10.1257/aer.p20161008
- Oostrom, J. K., & Born, M. P. (2014). Using cognitive pretesting to explore causes for ethnic differences on role-plays. *International Journal of Intercultural Relations*, 41, 138–149. doi:10.1016/j.ijintrel.2013.12.008
- Roe, R. A. (1983). *Grondslagen der personeelsselectie*. Assen: Van Gorcum.
- Rousseau, D. M., & Barends, E. G. R. (2011). Becoming an evidence-based HR practitioner. *Human Resource Management Journal*, 21(3), 221–235. doi:10.1111/j.1748-8583.2011.00173.x
- Rynes, S. L., & Barber, A. E. (1990). Applicant attraction strategies: An organizational perspective. *Academy of Management Review*, 15(2), 286–310.

doi:10.5465/AMR.1990.4308158

- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, 56(4), 302–318. doi:10.1037/0003-066X.56.4.302
- Schein, E. (1965). *Organizational psychology*. London: Prentice-Hall.
- Schleicher, D. J., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2010). If at first you don't succeed, try, try again: Understanding race, age, and gender differences in retesting score improvement. *Journal of Applied Psychology*, 95(4), 603–617. doi:10.1037/a0018920
- Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86(1), 162–173. doi:10.1037/0022-3514.86.1.162
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274. doi:10.1037/0033-2909.124.2.262
- Schmidt, F. L., Shaffer, J. A., & Oh, I.-S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, 61(4), 827–868. doi:10.1111/j.1744-6570.2008.00132.x
- Selor. (2017). *Studie redelijke aanpassingen. Het volledige onderzoek over redelijke aanpassingen bij testen in de selectiecontext van Selor*. Brussels, Belgium.
- Stone, E., Laitusis, C. C., & Cook, L. L. (2015). Increasing the accessibility of assessments through technology. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 217–234). New York: Routledge.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments (Synthesis Report 44)*. Minneapolis, MN.
- Verhaeghe, P.-P., & Van der Bracht, K. (2017). Praktijktesten: Van onderzoeksmethode naar beleidsinstrument tegen discriminatie? *Sociologos*, 38(1), 182–200.
- Vornholt, K., Villotti, P., Muschalla, B., Bauer, J., Colella, A., Zijlstra, F., ... Corbière, M. (2017). Disability and employment – overview and highlights. *European Journal of Work and Organizational Psychology*, 1–16. doi:10.1080/1359432X.2017.1387536
- 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(7), 1–19. doi:10.1080/133693183X.20105.11303279