

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

3-2011

Instruments for Personnel Selection in the 21st Century: Research and Practice [in Dutch]

Filip LIEVENS

Singapore Management University, filiplievens@smu.edu.sg

Britt DE SOETE

Ghent University

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the [Human Resources Management Commons](#), and the [Organizational Behavior and Theory Commons](#)

Citation

LIEVENS, Filip and DE SOETE, Britt. Instruments for Personnel Selection in the 21st Century: Research and Practice [in Dutch]. (2011). *Gedrag en Organisatie*. 24, (1), 18-42. Research Collection Lee Kong Chian School Of Business.

Available at: https://ink.library.smu.edu.sg/lkcsb_research/5713

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Instrumenten om Personeel te Selecteren in de 21ste Eeuw:

Onderzoek en Praktijk

*Filip Lievens & Britt De Soete**

Dit artikel beoogt een evidence-based overzicht te geven van innovatieve selectietechnieken die de laatste jaren zijn ontwikkeld om nieuwe en verfrissende antwoorden te geven op kernvragen in de hedendaagse selectiepraktijk. Ook beogen wij tekorten aan te stippen in de huidige kennis over deze technieken om op deze manier toekomstig onderzoek aan te sporen. Deze bijdrage laat zien dat nieuwe selectie-instrumenten als conditionele redeneertesten, integriteitstesten, en impliciete associatietesten mogelijkheden bieden voor de meting van waardegebonden competenties en maladaptieve trekken. Verder laten gecontextualiseerde persoonlijkheidsvragenlijsten, businessgerelateerde intelligentietesten, situational judgment tests, en zogenaamde ‘serious games’ toe om de selectieprocedures op te smukken met een aantrekkelijk imago, hetgeen mogelijkheden biedt voor de marketing van de organisatie als werkgever. Ten slotte constateren we dat zowel de selectiepraktijk als -wetenschap verschillende grote uitdagingen te wachten staan. Voorbeelden hiervan zijn de selectie van divers én competent personeel en het betrouwbaar en valide gebruik van ongesuperviseerde Internet testing.

1 Inleiding

De selectie van personeel vormt een van de oudste domeinen in de Arbeids- en Organisationspsychologie. Door deze lange geschiedenis krijgt personeelsselectie soms het etiket van ‘platgetreden en saai domein’ of van ‘niets nieuws onder de zon’. Toegegeven, de laatste honderd jaar nam onze kennis over de betrouwbaarheid en validiteit van cognitieve vaardigheidstests, persoonlijkheidsvragenlijsten, selectie-interviews, en assessment center oefeningen al met rasse schreden toe (Schmidt & Hunter, 1998). Dit maakt het voor wetenschappers en practici niet altijd gemakkelijk om vernieuwende bijdragen omtrent deze traditionele selectietechnieken te leveren.

Toch zijn er de laatste jaren een aantal interessante nieuwe selectieprocedures ontwikkeld, die inspelen op de behoeften van organisaties in de 21ste eeuw. Als start van deze reeks over selectie en rekrutering beogen wij deze nieuwe selectieprocedures te belichten. In het bijzonder geven wij een evidence-based overzicht

* Filip Lievens en Britt De Soete zijn verbonden aan de Universiteit Gent.

van deze nieuwe selectie-instrumenten. Ook stippen wij tekorten aan in de huidige kennis over deze technieken om op deze manier richting te geven aan toekomstig onderzoek.

Zoals aangehaald, ligt de klemtoon van dit artikel op nieuwe selectieprocedures. Bijgevolg willen we met opzet een ruimere invalshoek hanteren dan ons louter te focussen op de traditionele cognitieve vaardigheidstests, persoonlijkheidsvragenlijsten, selectie-interviews, en assessment center oefeningen. Laatstgenoemde technieken worden enkel vermeld indien ze een gedaanteverwisseling ondergingen (dat wil zeggen: grondig zijn aangepast in functie van de huidige behoeften in organisaties). Vermits de focus in dit artikel ligt op de specifieke selectieprocedures (of selectie-instrumenten), krijgt het ruimere selectieproces dan ook slechts zijdelings aandacht.

Onze bespreking van nieuwe selectieprocedures is opgebouwd rond vier vragen waarmee organisaties in 2011 geconfronteerd worden in het domein van personeelsselectie. Allereerst bespreken we hoe men selectieprocedures aantrekkelijker kan maken voor sollicitanten. Deze trend speelt in op het toegenomen belang voor organisaties om zich als werkgevers te onderscheiden van de concurrenten op de arbeidsmarkt. Een tweede vraag speelt in op de behoeften van organisaties om naar aanleiding van recente bedrijfsschandalen selectieprocedures tevens in te zetten om de 'duistere kant' van sollicitanten te kunnen inschatten. Een derde vraag betreft de efficiëntie van selectieprocedures en meer bepaald het gebruik van internettechnologie om sneller en goedkoper een grote groep sollicitanten te kunnen testen. Ten vierde en ten laatste geven we een overzicht van selectieprocedures die organisaties kunnen toelaten een competent én divers personeelsbestand te selecteren. Tot slot dient opgemerkt te worden dat deze lijst van vier vragen niet pretendeert om alle hedendaagse HR-kwesties te behandelen. Op grond van onze jarenlange samenwerking met organisaties zijn we echter van mening dat onderstaande topics een antwoord kunnen bieden op enkele van de meest prangende vragen in het huidige selectiedomein.

1.1 Hoe kunnen organisaties selectieprocedures aantrekkelijker maken?

Sollicitanten beschouwen het selectieproces en de selectieprocedures doorgaans als één van de 'visitekaartjes' van organisaties. Van hun kant gaan organisaties eveneens steeds meer aandacht besteden aan het aantrekkelijk maken van selectieprocedures als onderdeel van hun 'employer branding' (d.w.z. intern en extern promoten van de organisatie als aantrekkelijke werkgever, Lievens, 2007). Empirisch onderzoek bevestigt deze link tussen percepties over de selectieprocedure en de aantrekkelijkheid van de organisatie. Onderzoek wees bijvoorbeeld uit dat sollicitanten die zich een positief beeld hebben gevormd over de selectieprocedure aangaven sneller een loonaanbod van de organisatie te aanvaarden en de organisatie meer te zullen aanprijzen in hun omgeving (Hausknecht, Day & Thomas, 2004).

Voorgaand onderzoek legde zich voornamelijk toe op de percepties over selectieprocedures in het algemeen. Deze studies toonden aan dat kandidaten doorgaans de voorkeur gaven aan interviews en 'sample-based' testen boven cognitieve vaardigheidstests, persoonlijkheidsvragenlijsten en biodata (o.a. Hausknecht et al.,

2004; Iles & Mabey, 1993). Ook de determinanten van deze sollicitantenpercepties werden reeds uitvoerig onderzocht. Hierbij kwam naar voren dat de 'functie-relevantie' ('job-relatedness') van de selectieprocedure de belangrijkste beïnvloedende factor vormt van sollicitantenpercepties (Hausknecht et al., 2004; Ryan & Tippins, 2004). Om deze reden poogt men recentelijk deze inzichten te gebruiken om bestaande selectie-instrumenten aan te passen of om volledig nieuwe selectie-instrumenten te ontwerpen.

Zo wordt er op heden volop geëxperimenteerd met het gebruik van contextualisatie als methode om de functie-relevantie en aantrekkelijkheid van de selectie-instrumenten te verhogen. Contextualisatie verwijst naar het gebruik van functie-specifieke items in plaats van generieke items en lijkt alvast een veelbelovende techniek voor sommige selectieprocedures die door sollicitanten vaak als minder positief worden ervaren. Een voorbeeld is het gebruik van businessgerelateerde cognitieve vaardigheidstesten. De functie-relevantie in deze businessgerelateerde vaardigheidstesten wordt bevorderd door bij de itempresentatie gebruik te maken van realistische grafieken, tabellen of functiespecifieke tekstpassages. Hatrup, Schmitt en Landis (1992) erkenden dat sollicitanten doorgaans meer gemotiveerd zijn om te presteren op cognitieve vaardigheidstesten met een hoge face-validiteit en vergeleken daartoe traditionele cognitieve vaardigheidstesten met een gecontextualiseerde variant. Ze stelden vast dat beide vaardigheidstesten dezelfde onderliggende constructen maten, wat impliceert dat businessgerelateerde vaardigheidstesten als waardige vervanger kunnen fungeren van de traditionele vaardigheidstesten. Vermoedelijk heeft het contextualiseren van vaardigheidstesten eveneens positieve effecten op de voorspellende validiteit en de sollicitantenpercepties. Bijkomend onderzoek is echter nodig om deze veronderstellingen te bevestigen.

De contextualisatietechniek werd eveneens toegepast bij persoonlijkheidsvragenlijsten (Bing, Whanger, Davison & VanHook, 2004; Lievens, De Corte & Schollaert, 2008; Schmit, Ryan, Stierwalt & Powell, 1995). Onderzoek toonde aan dat verschillende respondenten verschillende referentieraamwerken hanteren bij het invullen van een persoonlijkheidsvragenlijst, waardoor de bruikbaarheid van de testresultaten in twijfel kan getrokken worden. Op een item als 'Ik besteed aandacht aan details' zullen de antwoorden van sollicitanten verschillen naargelang het gebruikte referentieraamwerk. Bijgevolg zullen sommige respondenten dit item beantwoorden in het kader van een sociale gebeurtenis, anderen in het kader van een werkcontext, en nog anderen in het kader van een algemene situatie. Om deze reden verschaft men in gecontextualiseerde persoonlijkheidsvragenlijsten de respondent een 'frame-of-reference' (bijvoorbeeld door het toevoegen van een contextlabel) waardoor iedere respondent hetzelfde referentieraamwerk gebruikt bij het beantwoorden van de items. Het voorgaande item zal er in een gecontextualiseerde persoonlijkheidsvragenlijst als volgt uitzien: 'Ik besteed aandacht aan details op het werk.' Er wordt geopperd dat de ambiguïteit van de doorgaans gebruikte niet-gecontextualiseerde items één van de redenen is voor de eerder lage voorspellende validiteit van persoonlijkheidsvragenlijsten (Bing et al., 2004; Schmit et al., 1995). Vergeleken met hun traditionele tegenhanger constateert men bij gecontextualiseerde vragenlijsten een verhoogde betrouwbaarheid

en voorspellende validiteit (Bing et al., 2004; Lievens et al., 2008; Schmit et al., 1995). Ook hier is echter aanvullend veldonderzoek noodzakelijk om na te gaan of de veronderstelling dat gecontextualiseerde vragenlijsten een positief effect hebben op sollicitantenreacties wel degelijk bevestigd wordt. Een andere vraag is hoever deze contextualisatie kan en moet worden doorgetrokken. Tot op heden werden voornamelijk lichte vormen van contextualisatie gebruikt in selectie-instrumenten (zoals het toevoegen van het contextlabel 'op het werk' in bovenstaand voorbeelditem). Men kan nog een stuk verder gaan in het contextualiseren van instrumenten, bijvoorbeeld door het ontwikkelen van vragenlijsten met zeer specifieke en gedetailleerde items afgestemd op specifieke sollicitantenpools. Men kan zich echter afvragen in hoeverre de bijkomende kosten en moeite om deze selectie-instrumenten te contextualiseren zich vertalen in een verbetering van de psychometrische eigenschappen van de test. Zo kan men vermoeden dat er op een bepaald niveau van contextualisatie een plafondeffect optreedt met betrekking tot voorspellende validiteit. Bijkomend onderzoek is nodig om een licht te werpen op de mogelijkheden en grenzen van deze contextualisatietechniek.

Andere selectieprocedures die meestal hoog scoren op realisme en functie relevantie zijn work samples en assessment centers. Om deze reden worden ze ook 'high-fidelity' simulaties genoemd. De kandidaat wordt hierbij geconfronteerd met een realistische functiesituatie en dient vervolgens onmiddellijk te reageren met de juiste handelwijze. Een nadeel is echter dat deze simulaties relatief duur zijn en vooral bij kleine groepen kandidaten (enkel in de eindfase van een selectie) kunnen worden afgenomen. Daarom heeft men de laatste tien jaar gezocht naar alternatieve selectie-instrumenten, die ook een redelijke graad van realisme en functie relevantie bezitten maar wel bij grote groepen sollicitanten (bij preselecties van grootschalige selectieprocedures) kunnen worden ingezet. In dit kader vormen Situational Judgment Tests (SJT's) het meest bekende alternatief om de selectieprocedure aantrekkelijker te maken door in te spelen op functie relevantie. In SJT's worden aan de sollicitant werkgerelateerde situaties gepresenteerd, waarna de sollicitant de meest passende handelwijze dient te selecteren uit een aantal aangeboden responsalternatieven (McDaniel, Morgeson, Finnegan, Campio & Braverman, 2001; Lievens, Peeters & Schollaert, 2007; Lievens & Schollaert, 2008). SJT's waarbij zowel de situaties als de antwoordalternatieven in geschreven vorm worden gepresenteerd, worden ook wel 'low-fidelity' simulaties genoemd. Niettegenstaande het feit dat SJT's in staat zijn om een veelheid aan competenties van de sollicitant in te schatten, bewijzen ze in de praktijk voornamelijk hun nut bij het bepalen van interpersoonlijke, team- en leiderschapscompetenties (Christian, Edwards & Bradley, 2010).

Het afgelopen decennium werd intensief onderzoek gevoerd naar het gebruik van SJT's als nieuw selectie-instrument. De empirische evidentie is over het algemeen positief. Zo constateerde men een behoorlijke voorspellende validiteit voor SJT's (Christian et al., 2010; McDaniel et al., 2001; McDaniel, Hartman, Whetzel & Grubb, 2007). Dit geldt zowel voor SJT's met gedragsgeoriënteerde instructies (bijv. 'Wat zou je meest waarschijnlijk doen?') als SJT's met kennisgeoriënteerde instructies (bijv. 'Wat is volgens jou het beste antwoord?'). Procedurele kennis en 'implicit trait policies' werden naar voren geschoven als twee mogelijke verklarin-

gen voor de goede validiteit van SJT's (Motowidlo, Hooper & Jackson, 2006). De zogenaamde 'implicit trait policy' verklaring poneert dat de persoonlijkheid van individuen vorm geeft aan hun oordelen aangaande de effectiviteit van gedragingen (antwoordalternatieven per SJT-item) die meer of minder een uiting zijn van de persoonlijkheidstrek in kwestie. Op deze manier wordt het mogelijk om afleidingen te maken over iemands persoonlijkheid op basis van de effectiviteitsoordelen van dit individu over verschillende antwoordalternatieven. Volgens sommigen is een bijkomende sterkte van SJT's hun vermogen om bijkomende variantie te verklaren boven op de reeds verklaarde variantie door andere selectie-instrumenten (incrementele validiteit). Zo rapporteren Clevenger, Pereira, Wiechmann, Schmitt en Harvey (2001) een significante incrementele validiteit van SJT's over een combinatie van cognitieve vaardigheid, jobervaring, jobkennis, en de persoonlijkheidstrek 'Consciëntieusheid'. Ook McDaniel et al. (2007) stelden aan de hand van hun meta-analyse vast dat SJT's een incrementele validiteit van .01 tot .02 vertonen boven op de combinatie van een cognitieve vaardigheidstest en een Big Five persoonlijkheidsvragenlijst. Verder suggereren de onderzoeksresultaten dat de SJT als een nuttige aanvulling in het selectieproces kan gebruikt worden naast een cognitieve vaardigheidstest of een persoonlijkheidstest afzonderlijk, aangezien SJT's hier respectievelijk 3-5% en 6-7% unieke bijkomende variantie verklaren (McDaniel et al., 2007). Tot slot toont onderzoek tevens aan dat SJT's minder subgroepverschillen genereren tussen leden van etnische minderheids- en meerderheidsgroepen dan traditionele selectietests (Whetzel, McDaniel & Nguyen, 2008).

Een andere troef van SJT's is hun hoge face-validiteit. Dit is vooral zo wanneer bij SJT's gebruik gemaakt wordt van video- en multimediamateriaal. Ook cartoons worden voor deze doeleinden gebruikt. Dergelijke SJT's met video- en multimediatechnologie worden ook wel 'medium-fidelity' simulaties genoemd. Onderzoek bevestigde dat het gebruik van SJT's met videofragmenten (Chan & Schmitt, 1997), multimedia-apparatuur (Richman-Hirsch, Olson-Buchanan & Drasgow, 2000) en interactiviteit (Kanning, Grewe, Hollenberg & Hadouch, 2006) door de sollicitanten als het meest aangenaam en face-valide werd ervaren. Ook blijken deze SJT's een hogere voorspellende validiteit te bezitten, hetgeen hun hogere ontwikkelingskosten kan compenseren (Lievens & Sackett, 2006). Daarnaast wordt ook geëxperimenteerd met webcam-SJT's (Meltzer, 1995; Oostrom, Born, Serlie & Van der Molen, 2010). In deze selectie-instrumenten wordt aan de sollicitant een multimediale SJT-itemstam gepresenteerd zonder dat hierop de traditionele multiplechoice (d.w.z. 'close-ended') aanbieding van mogelijke antwoordopties volgt. Na het aanbieden van het SJT-item dient de sollicitant zonder uitstel zelf (d.w.z. 'open-ended' of 'constructed-response') te reageren op de aangeboden functiegerelateerde situatie. Het aldus via een webcam verzamelde beeldmateriaal wordt nadien beoordeeld door getrainde assessoren. Dergelijke hybride selectie-instrumenten als de webcam-SJT tonen aan dat de grenzen tussen SJT's en assessment centers steeds meer vervagen in functie van de huidige organisatiebehoeften.

Een laatste ontwikkeling in het selectiedomein om selectieprocedures meer aantrekkelijk te maken voor sollicitanten bestaat uit het gebruik van 'serious games'

of 'virtual reality' (Reynolds & Weiner, 2009). In deze virtuele realiteitsomgeving wordt computertechnologie gehanteerd om gebruikers onder te dompelen in een virtuele wereld (Aguinis, Henle & Beaty, 2001). Drie kenmerkende eigenschappen die virtual reality games onderscheiden van andere selectietechnieken zijn de navigatiemogelijkheden, een hoge mate van interactie, en mogelijkheid tot onderdompeling (Vince, 1998). Aguinis et al. (2001) opperen dat deze games mogelijk meer valide zijn dan traditionele selectietechnieken omwille van hun hoog realiteitsgehalte. Ook in SJT's experimenteert men tegenwoordig met het gebruik van virtuele personages om de face-validiteit van het instrument te verhogen (Fetzer, Tuzinski & Freeman, 2010). Onderzoek naar de meetpretentie en de validiteit van deze 'virtuele realiteit' selectietechnieken is voorsnog niet gebeurd.

1.2 Hoe kunnen organisaties de duistere kant van sollicitanten inschatten?

In de laatste tien jaar vormt het onderzoek naar de onderliggende structuur van het prestatiedomein één van de belangrijkste ontwikkelingen bij personeelselectie. Deze aandacht voor de onderliggende dimensies van succesvolle werkprestaties is noodzakelijk omdat selectieprocedures effectieve werkprestaties (het zogenaamde criterium) proberen te voorspellen. Algemeen wordt nu erkend dat het prestatiedomein bestaat uit drie grote componenten: taakgedrag (een cluster van gedragingen die nauw verband houden met de kernprocessen in de organisatie), contextueel gedrag (een cluster van gedragingen die bijdragen tot een goed psychologisch klimaat in de organisatie, ook wel 'extrarolgedrag' of 'organizational citizenship behavior' genoemd) en antiproductief gedrag (een cluster van gedragingen die opzettelijk gericht zijn tegen de doelstellingen van de organisatie, Rotundo & Sackett, 2002). Verder werd vastgesteld dat cognitieve vaardigheidstests vooral goede voorspellers zijn voor taakprestatie, terwijl persoonlijkheidsvragenlijsten dan weer goede voorspellers zijn voor contextueel gedrag (Van Scotter, Motowidlo & Cross, 2000). Vanuit wetenschappelijk oogpunt is echter minder bekend over de predictoren van de antiproductieve component van het werkprestatiedomein. Ook organisaties tonen belangstelling om de duistere kanten van werknemers ('dark side behavior') te kunnen voorspellen via selectie. Zo voelen steeds meer bedrijven de behoefte om eventuele maladaptieve trekken van hun potentiële medewerkers te 'detecteren' gedurende het selectieproces. Organisaties worden immers geconfronteerd met een hoge diefstalprevalentie, oneerlijke medewerkers, en contraproductief werkgedrag. Daarenboven vertonen organisaties interesse om integriteit of persoonlijke discipline in hun selectieproces te meten aangezien de functie waarvoor gesolliciteerd wordt een hoog niveau van discretie en rechtschapenheid vergt. Voorbeelden van dergelijke beroepen zijn geldkoeriers, politiemedewerkers of werknemers in nucleaire centrales. Tot slot hebben waardegebonden competenties (bijvoorbeeld integriteit, ethisch handelen) een vaste plaats verworven in de competentieraamwerken van de meeste organisaties, waardoor hun integratie in de selectieprocedure een logische volgende stap is.

Vanuit zowel wetenschappelijke als praktische hoek stelt zich dus de vraag of er selectieprocedures voorhanden zijn om antiproductief gedrag te kunnen voorspellen. Een hierbij aansluitende vraag is in welke mate er reeds wetenschappelijke

evidentie beschikbaar is aangaande de predictie van antiproductief werkgedrag. Een eerste methode die organisaties kunnen aanwenden om antiproductief gedrag en – in het bijzonder – de integriteit in te schatten bij hun sollicitanten of werknemers, is het gebruik van een ruime waaier aan integriteitstesten. Deze testen kunnen onderverdeeld worden in twee categorieën, met name de ‘openlijke of overte integriteitstesten’ en de ‘persoonlijkheidsgebaseerde integriteitstesten’ (Berry, Sackett & Wiemann, 2007). De eerste variant omvat testinstrumenten waarbij de sollicitant openlijk bevraagd wordt over zijn of haar attitudes aangaande diefstal, oneerlijkheid en overig contraproductief werkgedrag. Ook wordt via zelfrapportage nagegaan in welke mate de medewerker in het verleden zelf illegaal of contraproductief gedrag gesteld heeft (Wanek, 1999). Voorbeelditems kunnen zijn: ‘Ben je ervan overtuigd dat iedereen in bepaalde mate oneerlijk is?’ en ‘Vind je dat een werknemer die betrapt is op het bestellen van zijn werkgever de collega’s moet verklikken die hem hebben geholpen bij de diefstal?’

De tweede categorie, met name de ‘persoonlijkheidsgebaseerde integriteitstesten’, bevragen de eerlijkheid van de sollicitant minder onverholen en zijn sterk gelijkend aan traditionele persoonlijkheidsvragenlijsten. Deze instrumenten gaan verscheidene persoonskenmerken na die gerelateerd zijn aan het al dan niet stellen van contraproductief werkgedrag, zoals afhankelijkheid, sociale conformiteitszin, gewetensvolheid, en aanvaarding van autoriteit. De Hogan Personality Inventory (Hogan & Hogan, 2002) is een voorbeeld van een veelgebruikte persoonlijkheidsgebaseerde integriteitstest waarvan tevens een Nederlandstalige versie ontworpen is. De test bevat 205 items die laden op 6 werkgerelateerde schalen. De respondent dient na ieder item aan te stippen of hij/zij het eens of oneens is met de geponeerde uitspraak. Voorbeelditems zijn ‘Ik sta erop het respect te krijgen dat ik verdien’ of ‘Ik weet wie mijn vijanden zijn’. Op basis van zijn of haar antwoordpatroon wordt afgeleid of de sollicitant kenmerken vertoont van een afwijkend persoonlijkheidsprofiel dat nadelig zou kunnen zijn voor het welzijn van de organisatie en haar medewerkers.

Een derde benadering bestaat uit het gebruik van conditionele redeneertests. Zo ontwierp James de conditionele redeneertest om de justificatiemechanismen na te gaan die sollicitanten aanwenden om hun gedragingen te rationaliseren (James et al., 2005; LeBreton, Barksdale, Robin & James, 2007). Een voorbeeld van een item uit deze conditionele redeneertest is het volgende:

De helft van alle huwelijken mondt uit in een scheiding. Eén van de redenen voor het grote aantal scheidingen is dat scheiden vandaag zeer snel en gemakkelijk kan gebeuren. Als een koppel overeenstemming bereikt over hoe ze hun eigendom eerlijk kunnen verdelen, dienen ze enkel nog de juiste administratieve stappen af te handelen en deze naar de rechtbank te brengen en dient men geen beroep te doen op advocaten gedurende het scheidingsproces.

Welke van de volgende stellingen is de meest redelijke conclusie die men uit bovenstaande tekst kan trekken?

- a. Wanneer mensen trouwen, worden ze ouder.

- b. Koppels zouden mogelijk langer bij elkaar blijven, indien een scheiding meer moeite zou kosten.
- c. Mannen scheiden vaker dan vrouwen.
- d. Als één van de huwelijkspartners een advocaat inhuurt, is hij of zij niet van plan het spel eerlijk te spelen.

(LeBreton et al., 2007)

In een conditionele redeneertest presenteert men dus aan de kandidaat een item dat lijkt op een redeneertest, waardoor de kandidaat vermoedt dat het een vaardigheidstest betreft. Vervolgens worden vier antwoordalternatieven aangeboden en dient de kandidaat het antwoord te kiezen dat het meest logische gevolg is van het gepresenteerde item. Twee van de antwoordalternatieven zijn hierbij nonsens (in bovenstaand item zijn dit alternatief a en c). Twee andere antwoordalternatieven variëren in hun hostileitsgraad. Zo heeft alternatief d in dit voorbeeld een hoge hostileitsgraad en alternatief b een lage hostileitsgraad. Het kiezen van een meer of minder agressief antwoordalternatief over een groot aantal van dergelijke items geeft vervolgens een indicatie van de agressiviteit van de sollicitant en wordt gebruikt als voorspeller van toekomstig contraproductief gedrag.

Een vierde en laatste aanpak¹ betreft het gebruik van impliciete associatietesten (IAT). Een impliciete associatietest kan omschreven worden als een latente responstaak die tot doel heeft om de impliciete attitudes, stereotypen, zelfbeoordeling en het zelfconcept van de respondent te meten (Rudman, 2008). Testen binnen deze categorie zijn gestoeld op de veronderstelling dat onderliggende, latente attitudes een invloed kunnen hebben op de uitvoering van een gegeven taak. Wanneer men een taak uitoefent die consistent is met de eigen cognitieve associaties (bijvoorbeeld: vrouwen worden geassocieerd met 'warmte' en mannen worden geassocieerd met 'macht') dan zal men de taak vlot en efficiënt uitoefenen – Rudman (2008) vergelijkt het gemak waarmee dergelijke taken uitgeoefend worden met het spreken van de moedertaal. Echter, wanneer men wordt gevraagd om taken uit te oefenen die inconsistent zijn met de eigen impliciete attitudes (bijvoorbeeld: vrouwen worden geassocieerd met 'macht' en mannen worden geassocieerd met 'warmte') dan zullen deze onderliggende latente associaties in competitie treden met de taakvereisten, waardoor dit de taakuitoefening vertraagt en bemoeilijkt en waardoor meer fouten optreden – Rudman (2008) maakt hier de vergelijking met het spreken van een nieuwe, vreemde taal. Onderzoek naar impliciete attitudes binnen het selectiedomein vond voornamelijk plaats binnen de context van discriminatie bij selectiebeslissingen of tijdens sociale interacties (McConnel & Leibold, 2001; Ziegert & Hanges, 2005).

Wat zijn de resultaten van het empirisch onderzoek over deze vier methoden om de duistere kanten van sollicitanten te meten? Onderzoek wees uit dat zowel de overte integriteitstesten, de persoonlijkheidsgebaseerde integriteitstesten als de conditionele redeneertest een meer dan behoorlijke predictieve validiteit hebben (Berry et al., 2007; James et al., 2005; Ones, Viswesvaran & Schmidt, 1993). Uit de meta-analyse van Ones et al. (1993) kwam naar voren dat kandidaten die hoog scoren op integriteitstesten niet alleen minder contraproductief werkgedrag stel-

len, maar zich daarenboven ook productiever gedragen. Een tweede voordeel betreft de incrementele validiteit. Vermits integriteitstesten en cognitieve vaardigheidstesten vrijwel ongecorreleerd zijn, verklaren beide instrumenten unieke variantie (Ones et al., 1993). Wat betreft het gebruik van integriteitstesten bij minderheidsgroepen, constateerden Ones en Viswesvaran (1998) dat de raciale subgroepverschillen bij overte integriteitstesten verwaarloosbaar zijn. Bijkomend onderzoek is nodig om na te gaan of dit voordeel eveneens geldt voor persoonlijkheidsgebaseerde integriteitstesten, conditionele redeneertesten en impliciete associatietests.

Een belangrijk vraagstuk in verband met het gebruik van integriteitstesten is 'fakability': de mogelijkheid om de test al dan niet te vervalsen door intentionele vertekening van de testantwoorden om zo een meer aanvaardbare score te bekomen. Niettegenstaande sollicitanten vanzelfsprekend de mogelijkheid hebben om hun score op zowel openlijke als persoonlijkheidsgebaseerde integriteitstesten artificieel op te krikken door het 'faken' van hun antwoorden, is er nog geen zekerheid dat dit wel degelijk gebeurt (Berry et al., 2007). Men vermoedt dat de testvertekening meer voorkomt in openlijke integriteitstesten dan in persoonlijkheidsgebaseerde integriteitstesten, maar uitgebreider onderzoek is noodzakelijk om deze veronderstellingen te bevestigen. Men heeft wel reeds vastgesteld dat conditionele redeneertesten beduidend minder vatbaar zijn voor 'faking'. Deze bevinding geldt weliswaar uitsluitend indien men de kandidaten niet informeert over het doel van de test. Indien men de sollicitanten wel verduidelijkt dat het gaat om een integriteitsmeting, constateert men een significante vertekening (LeBreton et al., 2007).

Tot op heden zijn de onderzoeksresultaten over impliciete associatietesten uiteenlopend en is er sprake van verdeeldheid over de bruikbaarheid en betrouwbaarheid van dit instrument. 'Critici' zijn van mening dat IAT-onderzoekers al te gemakkelijk uitgaan van de vermeende link tussen het bezitten van impliciete attitudes en het vertonen van (veeleer negatief) werkgedrag (zonder dat hiervoor voldoende wetenschappelijke evidentie voorhanden is), dat men ten onrechte veronderstelt dat onderzoeksresultaten in een laboratorium gegeneraliseerd kunnen worden naar de werkcontext, en dat te sterke conclusies worden geformuleerd op basis van de onderzoeksresultaten (zo is het merendeel van de blanke Amerikanen die een IAT ondergingen, gediagnosticeerd als 'anti' ten aanzien van zwarte Amerikanen – Blanton, Klick & Mitchell, 2009; Landy, 2008a). Om hun kritiek te staven, heranalyseerden Blanton et al. (2009) de onderzoeksresultaten van twee toonaangevende IAT-studies in het selectiedomein (McConnel & Leibold, 2001; Ziegert & Hanges, 2005). Ze kwamen tot de conclusie dat de vermeende robuuste relatie tussen IAT-scores en discriminatief gedrag niet teruggevonden wordt en dat bijgevolg de predictieve validiteit en de implicaties die geuit worden op basis van voorgaand IAT-onderzoek in twijfel kunnen getrokken worden. Blanton et al. (2009) pleiten dan ook voor systematisch veldonderzoek naar de relatie tussen impliciete attitudes en het vertonen van werkgedrag, waarbij toekomstige IAT-studies dienen te controleren voor het effect van expliciete opvattingen en attitudes. Ook Landy (2008b) formuleerde enkele suggesties voor toekomstig onderzoek binnen het IAT-domein. In eerste instantie opperde hij dat de effecten van

individuatie (d.w.z. het verstrekken van individuerende info) op impliciete attitudes dienen onderzocht te worden. Daarnaast pleitte hij voor meer aandacht voor de invloed van individuele, contextuele en organisatieverschillen op impliciete attitudes en stereotypering.

Ter verdediging weerleggen aanhangers van de IAT deze kritieken (o.a. Greenwald, 2008; Hanges & Ziegert, 2008; Rudman, 2008) door onder andere de wetenschappelijke waarde van laboratoriumonderzoek, de generaliseerbaarheid van het onderzoeksdesign naar de eigenlijke werkcontext, en de eenzijdige rapportering van IAT-studies door critici aan te halen. Daarenboven toonde onderzoek aan dat IAT's beschikken over robuuste psychometrische eigenschappen en weinig tot geen mogelijkheid bieden tot faking (Nosek, Greenwald & Banaji, 2007). In hun recente overzichtsstudie op basis van 122 onderzoeksrapporten identificeerden Greenwald, Poehlman, Uhlmann en Banaji (2009) de IAT bovendien als een betrouwbare predictor voor tal van gedragingen, oordelen en fysiologische metingen. Daarnaast stelden ze een hogere predictieve validiteit vast voor de IAT dan voor zelfrapporteringmethoden bij de voorspelling van discriminerende gedragingen op basis van huidskleur.

Concluderend kunnen we stellen dat tal van selectie-instrumenten reeds beschikbaar zijn voor de inschatting van de 'dark side' van sollicitanten of werknemers. Integriteitstesten en conditionele redeneertesten vertonen doorgaans behoorlijke psychometrische eigenschappen. Over de geschiktheid van IAT zijn wetenschappers meer verdeeld. Men kan stellen dat het gebruik van IAT als voorspeller van contraproductief gedrag op heden nog steeds in zijn kinderschoenen staat en bijgevolg meer onderzoeks aandacht verdient.

1.3 Hoe kunnen organisaties de efficiëntie van selectieprocedures verhogen?

Sinds een aantal decennia hebben nieuwe technologieën het selectiedomein veroverd omdat organisaties technologie als hét middel bij uitstek zien om het selectieproces en de selectieprocedures efficiënter (lees: sneller en goedkoper) te laten verlopen. Terwijl de eerste generatie computergebaseerde testen uitsluitend bestond uit traditionele 'pen-en-papier' testen die naar een computerversie vertaald werden, gebruikten organisaties in latere generaties de pc tevens om een meerwaarde te geven aan selectie-instrumenten. Het gebruik van adaptieve testen en multimediatesten zijn de bekendste voorbeelden van dit 'computer-enhanced' testing (Bartram, 2008; McBride, 1998).

De laatste jaren heeft echter vooral het 'unproctored Internet testing' (UIT – ongesuperviseerde internet testing) zich als de nieuwste generatie van pc-testen gemanifesteerd. UIT verwijst naar het proces waarin geselecteerd wordt aan de hand van internettesten waarbij de respondent de test invult zonder dat er sprake is van een traditionele menselijke 'proctor' of testleider (Tippins, 2009a). Het gebruik van UIT gaat gepaard met tal van efficiëntievoordelen (Bartram, 2008; Tippins et al., 2006). Met behulp van UIT kunnen sollicitanten de internettesten invullen waar en wanneer ze willen. Hierdoor worden ook de verplaatsings-, personeels- en locatiekosten tot nul herleid. Ook zorgt het internet voor een consistente afname van de tests. Dankzij de gebruikte internettechnologie gebeurt het item- en gegevensbeheer tevens gecentraliseerd, snel en efficiënt.

Vermits vele managers vrezen competente kandidaten te verliezen aan de concurrentie omwille van logge selectieprocessen, vormt de snelheid van UIT en de gereduceerde rekruteringscyclus een belangrijke troef voor de praktijk. Ten slotte zijn ook veel managers van mening dat het 'high-tech imago' van UIT een positieve invloed heeft op het organisatie-imago en bijgevolg op de attractiviteit van de organisatie bij sollicitanten.

Voor persoonlijkheidsvragenlijsten en vragenlijsten aangaande de persoon-organisatie fit wordt UIT reeds veelvuldig gebruikt. Echter, het gebruik van UIT in het kader van het online afnemen van cognitieve vaardigheidstesten lijkt minder voor de hand liggend omwille van de mogelijke risico's die hiermee gepaard gaan (Bartram, 2008; Burke, 2009; Tippins, 2009a). Door de ongestandaardiseerde test-omgeving vrezen velen voor de betrouwbaarheid van testen die via UIT worden afgenomen. Daarnaast merkt men ook tal van bedreigingen op voor de validiteit van de afgenomen test (Stanton, 1999). Een eerste bedreiging voor de validiteit betreft de testveiligheid. Omwille van de ongesuperviseerde setting kunnen kandidaten de items memoriseren, neerschrijven en eventueel delen met anderen of verkopen op het internet. Vermits UIT-procedures doorlopen worden door grote sollicitantenaantallen is piraterij een niet te onderschatten gevaar voor de testveiligheid. Een tweede bedreiging voor de validiteit is testvervalsing, vermits de kandidaat de hulp kan inroepen van anderen tijdens het vervolledigen van de test of een beroep kan doen op de eerder vermelde illegale antwoordsleutels. Een derde validiteitsbedreiging bij UIT betreft de bezorgdheid over de identiteit van de sollicitant. Er is immers geen zekerheid dat de sollicitant zijn internettest zelf invult in plaats van dit aan iemand anders te vragen, om zo de kans op slagen te verhogen.

Ondanks het debat over de pro's en contra's van UIT-gebruik voor cognitieve vaardigheidstesten (Tippins, 2009a; Tippins et al., 2006), hebben de meeste organisaties in de VS reeds gekozen om cognitieve vaardigheidstesten via UIT aan te bieden. Zo schrijft Tippins (p. 4, 2009b): 'Voor vele selectiepsychologen en werkgevers heeft de UIT-trein het station verlaten. Men heeft UIT aanvaard als een efficiënte, kosteneffectieve oplossing voor het testen van aanzienlijke en geografisch verspreide sollicitantengroepen. De vraag luidt bijgevolg niet "Moeten we gebruik maken van UIT?" maar eerder "Wat is de beste manier om UIT te gebruiken?"

Dit citaat illustreert dat de kwestie of UIT-gebruik voor cognitieve vaardigheidstesten al dan niet geoorloofd is, niet langer beschouwd wordt als de belangrijkste vraag binnen het UIT-debat. De onderzoeksaandacht is verschoven naar de omstandigheden waaronder UIT-gebruik voor cognitieve vaardigheden kan toegepast worden. In lijn met Tippins' stelling dienen wetenschap en praktijk bijgevolg samen te zoeken naar oplossingen voor de problemen die gepaard gaan met UIT-gebruik en naar condities waarbij de risico's op lagere betrouwbaarheid en validiteit zo veel mogelijk geminimaliseerd worden.

Zoals reeds aangehaald, vormen de voornaamste problemen die zich voordoen bij UIT-gebruik en waarvoor een oplossing dient gezocht te worden, de mogelijkheid tot testvervalsing, de testveiligheid en de mogelijkheid tot identiteitsvervalsing. Deze bedreigingen voor de psychometrische eigenschappen van de test ontstaan

ten gevolge van het wegvallen van controle in de UIT-setting (Bartram, 2008). Om oneerlijk gedrag bij UIT te ontmoedigen werden reeds tal van suggesties geformuleerd.

Een eerste categorie ter optimalisatie van de testbescherming bij UIT betreft de mechanische strategieën (Cizek, 1999). Deze strategieën streven naar een verhoging van de testveiligheid door het hinderen en dwarsbomen van zogenaamde piraten of respondenten met oneerlijke bedoelingen. Verschillende strategieën binnen deze categorie werden reeds geïmplementeerd in organisaties die gebruik maken van UIT voor cognitieve vaardigheidstesten. Een voorbeeld hiervan omvat het voorzien van technische voorzorgsmaatregelen, bijvoorbeeld de onmogelijkheid om bepaalde functies op het toetsenbord of de computer te activeren gedurende de testafname (Foster, 2009). Een ander voorbeeld van een mechanische strategie bestaat erin om zogenaamde 'web patrols' op te richten waardoor men op regelmatige basis het internet scant op zoek naar sites die testinformatie, testitems of antwoorden (al dan niet tegen betaling) aanbieden (Burke, 2009).

Een zorgvuldige opmaak van de test (testdesign) is waarschijnlijk een van de meest beloftevolle mechanische strategieën. Het doel van testdesign binnen het kader van UIT bestaat uit het ontwikkelen van verschillende equivalente testen met dezelfde lengte waarbij de gebruiksfrequentie van ieder item zorgvuldig gecontroleerd wordt (Tippins et al., 2006). Om automatisch dergelijke testformats te ontwikkelen uit grote itempools zijn verschillende benaderingen voorhanden (Jodoin, Zenisky & Hambleton, 2006). Een eerste is het gebruik van 'linear on-the-fly-testing'; een techniek waarbij verschillende testen dynamisch (dat wil zeggen: tijdens de testafname) worden ontwikkeld op basis van een gekalibreerde itempool. In andere benaderingen worden items (zoals bij adaptief testen via de computer) of verzamelingen van items (zoals bij 'multistage testing') geselecteerd op het moment van testafname op basis van de huidige inschatting van de bekwaamheid van de kandidaat. Vanzelfsprekend dient men over een voldoende uitgebreide itempool te beschikken om testdesign als een succesvolle beveiligingsstrategie te kunnen hanteren.

Tot slot biedt ook het testformaat tal van beveiligingsmogelijkheden. Zo formuleerde Foster (2009) op basis van onderzoek een suggestie om de testbescherming te optimaliseren door wijzigingen in de aanbieding van de antwoordalternatieven. Terwijl de respondent in een traditionele test na ieder item alle responsalternatieven te zien krijgt en daaruit een keuze moet maken, worden de responsalternatieven in de opzet van Foster slechts één voor één in random volgorde gepresenteerd na het item. Na ieder gepresenteerd antwoordalternatief dient de respondent aan te geven of dit al dan niet het juiste antwoord op de vraag is tot hij (correct of incorrect) 'juist' geantwoord heeft of onterecht 'fout' geantwoord heeft. Op deze manier krijgt de kandidaat vrijwel nooit alle antwoordalternatieven te zien, waardoor volgens Foster de kansen op piraterij gevoelig dalen.

Een tweede categorie veiligheidsmaatregelen omvat de principiële strategieën (Cizek, 1999). Principiële strategieën beogen de intenties van de respondent om testfraude te begaan te reduceren door eerlijk antwoordgedrag te stimuleren. Binnen dit kader hebben tal van onderzoeken aangetoond dat het inbouwen van aansprakelijkheid in testprocedures geldt als een van de belangrijkste mechanismen

om respondenten te doen afzien van hun intentie om de antwoorden te vervalsen (Lerner & Tetlock, 1999). Het inbouwen van aansprakelijkheid kan op diverse manieren geoperationaliseerd worden, zoals de aanwezigheid van anderen, identificeerbaarheid en verifieerbaarheid (Lerner & Tetlock, 1999). De eerste twee tactieken, met name het voorzien van de aanwezigheid van anderen (de verwachting dat de testafnemer het gedrag van de respondent zal observeren) en identificeerbaarheid (de verwachting dat de scores van iedere respondent persoonlijk aan hem of haar gekoppeld zullen worden) zijn strategieën die voornamelijk succesvol bleken bij het stimuleren van eerlijk testgedrag in gesuperviseerde testomgevingen ('proctored testing'). Vermits deze strategieën bij UIT moeilijker te implementeren zijn, focust men in ongesuperviseerde settings voornamelijk op verifieerbaarheid (de verwachting dat de testprestatie opnieuw zal gemeten worden aan de hand van een andere methode en dat hieraan mogelijke consequenties verbonden zijn) om de aansprakelijkheid van de respondenten te verhogen. Deze aansprakelijkheidsdimensie wordt tevens de verwachting van validatie genoemd, vermits men aanneemt dat de zelfgerapporteerde informatie vergeleken zal worden met een andere prestatiemeting (Farh & Werbel, 1986). Zoals wordt aangeraden in de 'International Guidelines on Computer-Based and Internet-Delivered Testing' (Guideline 45.3, ITC, 2006) kan men respondenten informeren over de mogelijkheid dat hun scores op de UIT zullen geverifieerd worden door een andere test (verificatietest) in een gesuperviseerde testomgeving. Sterk verschillende scores tussen de 'proctored' en 'unproctored' versie van de test kunnen erop wijzen dat de kandidaat hulp kreeg bij het invullen van de UIT of de test door iemand anders liet invullen. In dat geval dient te kandidaat de 'unproctored' test opnieuw te maken onder supervisie.

Tot op heden is onderzoek naar de effecten van bovenstaande mechanische en principiële veiligheidsmaatregelen echter beperkt (Arthur, Glaze, Villado & Taylor, 2010; Nye, Do, Drasgow & Fine, 2008). In schril contrast met de populariteit van UIT in de VS, is onderzoek over UIT in actuele selectiesituaties vooralsnog heel schaars. Niet alleen de *preventie* van testbedrog genoot tot op heden weinig onderzoeksbelangstelling; ook de *detectie* van testbedrog is een vraagstuk dat nog niet opgelost is. Tot nu toe worden voornamelijk twee benaderingen gebruikt om testbedrog te detecteren. De eerste statistische methode vergelijkt de ongesuperviseerde en de gesuperviseerde testscore van de respondent (Nye et al., 2008). Er is vervolgens sprake van een afwijkend antwoordpatroon indien de testscores van de respondent in de gesuperviseerde setting significant lager liggen dan hun UIT-scores. Deze eerste methode bouwt dus voort op het verificatiemodel, waarbij de participanten die slagen voor de UIT-fase vervolgens een soortgelijke test in een gesuperviseerde setting dienen af te leggen. Een tweede methode om testbedrog te detecteren omvat het beroep doen op 'data forensics' – een benadering waarbij de antwoordpatronen van respondenten geanalyseerd worden en afwijkingen op dit patroon vastgesteld worden (Bartram, 2008; Foster, Maynes & Hunt, 2008). Bijgevolg worden respondenten die hun test te snel afleggen (omdat ze beschikken over een antwoordsleutel) of net te traag antwoorden (omdat ze de testinhoud kopiëren) geïdentificeerd. Hoewel beide statistische methodes vaak naar voor geschoven worden als systematische benaderingen om testbedrog te detecte-

ren, dient benadrukt te worden dat deze methodes in een operationele setting niet noodzakelijk een indicatie geven van testbedrog vermits ze onderhevig zijn aan type I fouten (foutief detecteren van testbedrog) en type II fouten (testbedrog niet als dusdanig detecteren), waardoor ze bij voorkeur in combinatie met andere indicatoren van testbedrog gebruikt worden (Cizek, 1999). Bijgevolg voorzien deze methodes geen absolute oplossing voor de detectieproblemen van de identificeerbaarheid van de respondent en/of assistentie bij het vervullen van de test. Dit laatste probleem kan in de toekomst mogelijk verholpen worden door het gebruik van webcamsupervisie of biometrische identificatie. Wat betreft deze laatste strategie adviseert Foster (2009) de implementatie van 'key stroke analysis', waarbij de respondent gedurende het invullen van de test geïdentificeerd wordt aan de hand van zijn of haar typpatroon.

Kortom, in recente discussies aangaande UIT verschoof de klemtoon van de vraag 'Is het gebruik van UIT van cognitieve vaardigheid geschikt in reële selectiesettings?' naar de vraag 'Hoe kunnen organisaties binnen deze settings de betrouwbaarheids- en validiteitsbedreigingen voor UIT maximaal reduceren?' (Tippins, 2009b). Een dergelijke verschuiving in de klemtoon is voornamelijk relevant voor organisaties die vandaag reeds geïnvesteerd hebben in grootschalige UIT-procedures van cognitieve vaardigheid. Daarnaast impliceert bovenstaande verschuiving eveneens een sleutelrol voor selectiepsychologen in het vormgeven van technologische oplossingen om zodoende zowel aan de noden van ondernemingen als aan de teststandaarden te voldoen. Een van de belangrijkste uitdagingen hierbij betreft het effect van UIT (met verificatietesting) op de validiteit van de test. Het verificatiemodel wordt immers vaak bekritiseerd omdat men mogelijk compromitterende UIT-scores gebruikt als criterium om de kandidaten uit te nodigen voor het volgende (gesuperviseerde) selectiestadium, waardoor bepaalde ongeselecteerde respondenten mogelijk foutief uit de boot vallen. Daarenboven leidt het koppelen van een verificatietest aan de ongesuperviseerde test tot een gevoelige stijging van de selectiekosten. Anderen stellen dat het screenen van grote kandidatenpools door middel van UIT van een valide predictor als cognitieve vaardigheid nog steeds te verkiezen is boven cv-screening. Men dient nog te onderzoeken welke scores (ongesuperviseerde score, gesuperviseerde score of beide) opgenomen worden als basis voor de selectiebeslissing en welke impact dit heeft op de validiteit. Daarnaast is het belangrijk om te onderzoeken welke selectieratio het best gebruikt wordt in UIT met het oog op het minimaliseren van valse positieven en valse negatieven.

1.4 Hoe kunnen organisaties een divers personeelsbestand selecteren?

Diversiteit is een topic dat de laatste jaren toenemende onderzoeks aandacht genoot en gevoelig aan populariteit won in de HR-wereld. Naast een wettelijk kader waaraan men moet voldoen aangaande de samenstelling van het personeelsbestand, hebben organisaties vaak sociale of praktische redenen om te opteren voor een divers personeelsbestand. Zo willen tal van bedrijven sollicitantengroepen aanspreken die vandaag nog onvoldoende benut worden (bijvoorbeeld: allochtone werknemers, werknemers met een handicap, oudere werknemers,...) en zodoende een verkwisting van talent vermijden. Daarnaast kiezen sommige

organisaties ook weloverwogen voor het tewerkstellen van leden uit minderheids-groepen vanuit hun sociaal bewustzijn of omdat ze streven naar een reflectie van de maatschappelijke diversiteit in hun personeelsbestand. Onderzoek wees tevens uit dat heterogene groepen een beter besluitvormingsproces vertonen dan homogene groepen (Sommers, 2006) en menig werkgever verwacht dat een divers personeelsbestand de creativiteit in de organisatie ten goede komt. Daarnaast heeft het selecteren van een divers personeelsbestand ook nog bijkomende voordelen wanneer er schaarste heerst op de arbeidsmarkt. Het aanspreken van niet-traditionele sollicitantengroepen biedt organisaties immers niet alleen de mogelijkheid om al hun openstaande vacatures in te vullen, maar heeft daarenboven ook economische voordelen omdat binnen deze sollicitantenpools minder concurrentiestrijd heerst dan op de traditionele markten (Rynes & Barber, 1990). Ten slotte constateert men ook een steeds meer divers wordend klantenbestand, waardoor een diverse staf niet alleen kan bijdragen aan de groei van het klantenbestand, maar tevens kan inspelen op de uiteenlopende noden van deze klanten.

Om deze wettelijke, maatschappelijke en bedrijfskundige redenen dienen de gehanteerde instrumenten en procedures in het kader van personeelsbeleid en -selectie geëvalueerd te worden op de hoeveelheid en de omvang van de eventuele subgroepverschillen die ze genereren. De grootte van de subgroepverschillen bepaalt of een selectie-instrument al dan niet 'adverse impact' teweegbrengt en bijgevolg mogelijk als discriminerend wordt beschouwd. Volgens de Amerikaanse 'Uniform Guidelines for Employee Selection Procedures' (UGESP; Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor & Department of Justice, 1978) wordt 'adverse impact' gedefinieerd als een substantieel afwijkende selectieratio bij het in dienst nemen, het promoveren of het nemen van andere werkgerelateerde beslissingen waarbij leden van minderheids-groepen benadeeld worden, waardoor deze groepen ondergerepresenteerd worden in de organisatie.

Het bevorderen van diversiteit gaat doorgaans gepaard met het vermijden van discriminatie en het reduceren van subgroepverschillen. Niettegenstaande reeds vele strategieën werden geformuleerd om dit te bewerkstelligen, blijkt het selecteren van een divers en competent personeelsbestand nog steeds een complexe uitdaging. Immers, hoewel de cognitieve vaardigheidstest kan beschouwd worden als één van de meest valide voorspellers van functieprestaties (Schmidt & Hunter, 1998), stelden tal van onderzoekers substantiële subgroepverschillen vast in de resultaten van verschillende vaardigheidstesten ten nadele van minderheids-groepen (Sackett, Schmitt, Ellingson & Kabin, 2001). De kernvraag in dit domein luidt bijgevolg of er alternatieve selectiestrategieën kunnen ontwikkeld worden die zowel weinig adverse impact genereren als over een behoorlijke validiteit beschikken. In een aanloop naar het beantwoorden van de diversiteit-validiteit vraag onderscheiden Ployhart en Holtz (2008) hierbij vijf categorieën. Een eerste categorie omvat het gebruik van selectie-instrumenten die kleinere subgroepverschillen genereren dan cognitieve vaardigheidstesten. Zo kan men gebruik maken van alternatieve meetinstrumenten zoals assessment centers, situational judgment tests en work samples (Chan & Schmitt, 1997; Pulakos, Schmitt & Chan, 1996; Schmitt & Mills, 2001). De verklaring voor de daling in subgroepverschillen bij

deze simulatieoefeningen ligt volgens velen bij de lagere leesvereisten, de positievere sollicitantenpercepties en de focus op niet-cognitieve competenties. Echter, onderzoekers wijzen erop dat de subgroepverschillen – hoewel gedaald – nog steeds aanzienlijk blijven (Dean, Roth & Bobko, 2008; Roth, Bobko, McFarland & Buster, 2008). Andere strategieën binnen deze categorie zijn het gebruik van schoolresultaten als een proxy variabele voor cognitieve vaardigheid of het opteren voor specifieke vaardigheidstesten (verbaal, numeriek,...) in plaats van algemene cognitieve vaardigheidstesten. Beide strategieën hebben kleine tot matige dalingen in subgroepverschillen tot gevolg (Berry, Gruys & Sackett, 2006; Hough, Oswald & Ployhart, 2001; Roth & Bobko, 2000).

Een tweede categorie strategieën met het oog op reductie van subgroepverschillen betreft het combineren of wegeven van testcores. Zo raden verschillende onderzoekers aan om zowel cognitieve als niet-cognitieve selectie-instrumenten of competenties op te nemen in de testbatterij (o.a. Sackett & Ellingson, 1997). Deze tactiek leidt niet alleen tot een reductie van de subgroepverschillen, maar komt eveneens de validiteit van het selectie-instrument ten goede en kan daarom bestempeld worden als een van de betere strategieën om het diversiteit-validiteit dilemma te benaderen (Sackett et al., 2001). Daarnaast behoort ook het toekennen van gewichten aan de verscheidene selectie-instrumenten of selectiecriteria tot de mogelijkheden. Echter, de effecten van deze laatste strategieën bleken tot op heden eerder beperkt (De Corte & Lievens, 2003; Hatrup, Rock & Scalia, 1997). Daarenboven wijzen Sackett et al. (2001) erop dat de doorslaggevende factor bij het toekennen van deze gewichten steeds de functievereisten dienen te zijn en dat dit proces niet louter mag gebeuren met het oog op de verhoopde reductie in subgroepverschillen. Een laatste strategie die binnen deze tweede categorie valt, is het toepassen van 'banding', waarbij men het volledige spectrum van scores onderverdeelt in 'banden' waarbinnen men geen onderscheid maakt tussen kandidaten die binnen een zelfde 'band' vallen. Deze techniek is gebaseerd op de veronderstelling dat kleine verschillen in testcores niet betekenisvol zijn omdat ze vallen binnen het interval van scores dat kan ontstaan ten gevolge van de meetfout. De grootte van de band wordt dus bepaald door de betrouwbaarheid van de test; waarna de selectie binnen een band gebeurt op basis van criteria die minder subgroepverschillen vertonen (Campion et al., 2001). Deze bandingstechniek is echter omstreden en in vele landen verboden. Bovendien kan 'banding' de testvaliditeit negatief beïnvloeden (Ployhart & Holtz, 2008).

Een derde categorie bestaat uit het reduceren van irrelevante testvariantie. Dit bekomt men door het beperken van veelal onnodige verbale vereisten van het selectie-instrument, wat doorgaans tot een aanzienlijke daling van de subgroepverschillen leidt. Een illustratie van deze tactiek is onder andere het gebruik van video-SJT's in plaats van geschreven SJT's (Chan & Schmitt, 1997). Ook 'interactive voice response' technologie kan binnen deze optiek gebruikt worden. Zo wees onderzoek uit dat het gebruik van een telefoon door de selectiepsycholoog om de selectievragen te communiceren en het gebruik van een telefoontoetsenbord door de sollicitant om het gekozen antwoordalternatief door te geven resulteerde in minimale subgroepverschillen (Van Iddekinge, Eidson, Kudisch & Goldblatt, 2003). Ook het gebruik van 'cultuurvrije' items – die geen enkele (culturele) sub-

groep bevoordelen – en ‘sensitivity panels’ – waarbij de panelleden beoordelen of de items geschikt en niet aanstootgevend zijn – worden als mogelijke strategieën naar voren geschoven. Noemenswaardige dalingen in subgroepverschillen ten gevolge van beide technieken werden vooralsnog niet opgetekend (Ployhart & Holtz, 2008). Hetzelfde geldt voor ‘differential item functioning’ (d.w.z. het identificeren en verwijderen van items die een systematisch onderscheid maken tussen subgroepen) en het wegnemen van tijdsbeperkingen gedurende de test (Sackett et al., 2001).

De vierde categorie omvat alle strategieën die een vorm van oefening toelaten in het selectieproces, zoals het organiseren van oriëntatieprogramma’s, het aanbieden van coaching of het voorzien van hertestmogelijkheden. Onderzoek constateerde opnieuw slechts beperkte en weinig consistente effecten ten gevolge van deze tactieken (Ployhart & Holtz, 2008; Sackett et al., 2001).

De vijfde en laatste categorie bestaat uit strategieën ter bevordering van de reacties van de deelnemende sollicitanten. Bijgevolg worden onder andere interventies opgesteld om leden van minderheidsgroepen te ontraden om uit de selectieprocedure te stappen; doorgaans met slechts beperkt succes wat betreft de reductie van subgroepverschillen in de uiteindelijke selectieresultaten (Tam, Murphy & Lyall, 2004). Daarnaast tracht men eveneens de percepties van de sollicitanten te optimaliseren door ingrepen in de selectieprocedure. Zo stelden Edwards en Arthur (2007) vast dat open vragen – waarop de respondent zelf een antwoord dient te formuleren – positiever worden onthaald bij minderheden en bijgevolg minder subgroepverschillen genereren dan multiplechoicevragen bij een kennis-test.

Hoewel bovenstaande strategieën reeds een stap in de goede richting zijn om het diversiteit-validiteit dilemma te beantwoorden, blijft het ontwikkelen van een instrument dat subgroepverschillen vermindert en daarenboven de voorspellende validiteit van de test garandeert een moeilijke evenwichtsoefening. Bijkomend onderzoek is noodzakelijk om zowel de kwantiteit als de kwaliteit van deze richtlijnen te optimaliseren en om te bepalen wat de potentiële meerwaarde van deze selectiemethodes is naast het gebruik van de traditionele cognitieve vaardigheidstest.

2 Een blik in de toekomst

De doelstelling van deze bijdrage bestond erin te belichten hoe de laatste jaren nieuwe selectietechnieken zijn ontwikkeld om nieuwe en verfrissende antwoorden te geven op kernvragen in de hedendaagse selectiepraktijk. Zo bieden nieuwe selectie-instrumenten als conditionele redeneertesten, integriteitstesten en impliciete associatietesten een mogelijk alternatief voor de meting van waardegebonden competenties en maladaptieve trekken. Empirisch onderzoek is nodig om de validiteit van sommige van deze technieken verder aan te tonen. Gecontextualiseerde persoonlijkheidsvragenlijsten, businessgerelateerde intelligentietests, diverse vormen van SJT’s, en serious games laten toe om het imago van de selectieprocedures aantrekkelijker te maken, wat opportuniteiten met zich meebrengt

betreffende de marketing van selectietechnieken en van de organisatie als werkgever. Ten slotte concluderen we dat zowel het werkveld als de wetenschap verschillende grote uitdagingen te wachten staan. Voorbeelden hiervan zijn de selectie van divers én competent personeel en het betrouwbaar en valide gebruik van UIT.

Betreffende toekomstig onderzoek vermoeden we dat studies binnen het domein van personeelsselectie zich vooral dienen te verdiepen langs drie assen. In eerste instantie dient gefocust te worden op het uitdiepen van het inhoudelijke aspect. Hoewel reeds grote vooruitgang geboekt is betreffende de voorspellers van taakgedrag en contextuele prestatie in organisaties, werd slechts weinig aandacht besteed aan andere inhoudelijke componenten van het criteriumdomein zoals adaptief werkgedrag (Pulakos, Arad, Donovan & Plamondon, 2000; Pulakos et al., 2002) en het eerder besproken antiproductief gedrag.

Een tweede as betreft het niveau waarop onderzoek wordt gevoerd. Tot op heden werden voornamelijk individuele resultaten (functieprestatie van de werknemers) bestudeerd. Indien het selectiedomein een rol wil spelen in strategisch HRM, dienen ook teamresultaten en organisatieresultaten onderzocht te worden als afhankelijke variabelen. Een voorbeeld van deze multi-level aanpak is het onderzoek van Ployhart, Weekley en Baughman (2006). Om inzicht te bekomen in de sterkte of specificiteit van de homogeniteit in persoonlijkheid van de geselecteerde individuen op de verschillende organisatieniveaus, deden Ployhart et al. een beroep op multi-levelonderzoek. Hun onderzoeksresultaten suggereren verschillende persoonlijkheidsstructuren al naargelang het organisatieniveau. Naast een vernieuwde focus op alle organisatieniveaus, dienen onderzoekers tevens aandacht te besteden aan een vergelijking van selectie op nationaal en internationaal niveau. Zo stelden Lievens, Harris, Van Keer en Bisqueret (2003) vast dat selectietechnieken die nuttig bleken op nationaal niveau niet steeds even bruikbaar zijn in een internationale selectiecontext. Hun onderzoek binnen het domein van cross-cultureel aanpassingsvermogen wees uit dat trainingssucces voorspeld werd door een combinatie van cognitieve vaardigheid, de persoonlijkheidstrek openheid, en assessment center beoordelingen van aanpassingsvermogen, teamwork en communicatievermogen. Op basis van hun onderzoeksresultaten suggereren ze om te opereren voor een verbreding van de selectieprocedure voor internationaal personeel aangezien deze vandaag doorgaans uitsluitend bestaat uit een inschatting van functiekennis en technische competenties.

Een derde en laatste as die als richtlijn voor toekomstig onderzoek kan functioneren betreft de tijdsdimensie. Een basisassumptie van het selectieparadigma is de relatieve stabiliteit van de 'functie', die als een standaard dient om de geschiktheid van sollicitanten te bepalen. Snel evoluerende organisatiestructuren (ten gevolge van samensmeltingen, afdankingen, teamwork of globalisatie) hebben geleid tot functie-instabiliteit en daarmee gepaard gaande nieuwe uitdagingen voor personeelsselectie (Kehoe, 2000). Er moet dus een inspanning geleverd worden om veranderingen in het criteriumconstruct op relatief korte termijn (eerste jaren) en op lange termijn (loopbaan) te voorspellen. Een mooi voorbeeld hiervan is het onderzoek van Stewart en Nandkeolyar (2006). Zij vergeleken inter-individuele en intra-individuele variatie in wekelijkse prestaties in een verkopers-

steekproef. Men stelde vast dat er grotere intra-individuele dan inter-individuele variatie in prestatie bestond en dat dit prestatieniveau daarenboven afhankelijk was van het niveau van aanpassingsvermogen van de verkoper. Op hun beurt beïnvloeden persoonlijkheidstrekken, zoals consciëntieusheid en openheid, de mate waarin een individu zich kan aanpassen aan veranderende omstandigheden en bijgevolg het prestatieniveau. Het dynamisch perspectief, waardoor deze studie gekenmerkt wordt, kan een vruchtbare inspiratiebron betekenen voor toekomstig onderzoek in het selectiedomein.

Praktijkbox

- In het kader van ‘employer branding’ trachten verscheidene organisaties hun selectie-imago te bevorderen door positieve sollicitantenreacties na te streven. Dit wordt onder meer bereikt door selectie-instrumenten te gebruiken die in hoge mate functie-relevant zijn, zoals assessment centers, work samples, SJT's en andere simulatievarianten. Ook het contextualiseren van vaardigheidstesten of persoonlijkheidsvragenlijsten lijkt binnen deze optiek een beloftevolle strategie.
- Tal van organisaties beogen integer gedrag binnen hun bedrijf te bevorderen en contraproductief gedrag te beperken. Dit is mogelijk door het selectieproces uit te breiden met integriteitstesten of conditionele redeneertesten, waarvan de psychometrische eigenschappen doorgaans veelbelovend zijn.
- Indien organisaties grote of geografisch verspreide sollicitantengroepen testen in de eerste fases van het selectieproces, kunnen ze hun procedures efficiënter laten verlopen door gebruik te maken van ‘ongesuperviseerde internet testing’. Met het oog op de betrouwbaarheid en validiteit van de test, dienen mechanische en principiële maatregelen getroffen te worden om testfraude te voorkomen en te detecteren.
- Organisaties die opteren voor een divers personeelsbestand kunnen gebruik maken van selectie-instrumenten die zo weinig mogelijk irrelevante variantie genereren en positieve sollicitantenpercepties opwekken zoals (multimediale) simulatietests. Daarnaast leidt het combineren van verscheidene cognitieve en niet-cognitieve selectie-instrumenten niet alleen tot minder subgroepverschillen maar ook tot een hogere predictieve validiteit.

Summary

Instruments for Personnel Selection in the 21st Century: Research and Practice
This article presents an evidence-based overview of innovative selection techniques that were developed in the last years to provide an answer to key questions concerning the contemporary practice of selection. In addition, we aim to point out gaps in the current knowledge about these novel techniques to guide future research. The article shows that innovative selection instruments such as condi-

onal reasoning tests, integrity tests, and implicit association tests might offer opportunities to measure value-based competencies, and maladaptive traits. Furthermore, contextualized personality questionnaires, business-related intelligence tests, situational judgment tests, and so-called serious games enable organizations to improve their image, which creates branding opportunities to promote the company as an attractive employer. Finally, we conclude by stating that practitioners and researchers face numerous selection challenges. Examples are the selection of a diverse as well as competent workforce and the reliable and valid implementation of unproctored Internet testing.

Keywords: Innovative selection techniques, Employer branding, Integrity testing, Internet testing, Diversity

Noot

- 1 Ook SJT's kunnen aangewend worden om integriteit van sollicitanten of medewerkers te meten. Pionierswerk op dit vlak werd verricht door Becker (2005), die een SJT ontwikkelde om integriteitgerelateerde werkuitkomsten te voorspellen.

Literatuur

- Aguinis, H., Henle, C.A. & Beaty, J.C. (2001). Virtual reality technology: A new tool for personnel selection. *International Journal of Selection and Assessment*, 9, 70-83.
- Arthur, W., Glaze, R.M., Villado, A.J. & Taylor, J.E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored Internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*, 18, 1-16.
- Ash, P. (1971). Screening employment applicants for attitudes toward theft. *Journal of Applied Psychology*, 55, 161-164.
- Bartram, D. (2008). The advantages and disadvantages of on-line testing. In S. Cartwright & C.L. Cooper, *The Oxford Handbook of Personnel Psychology* (pp. 234-260). Oxford: Oxford University Press.
- Becker, T.E. (2005). Development and validation of a situational judgment test of employee integrity. *International Journal of Selection and Assessment*, 13, 225-232.
- Berry, C.M., Gruys, M.L. & Sackett, P.R. (2006). Educational attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology*, 91, 696-705.
- Berry, C.M., Sackett, P.R. & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, 60, 271-301.
- Bing, M.N., Whanger, J.C., Davison, H.K. & VanHook, J.B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology*, 89, 150-157.
- Blanton, H., Klick, J. & Mitchell, G. (2009). Strong claims and weak evidence: reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, 94, 567-582.
- Burke, E. (2009). Preserving the integrity of online testing. *Industrial and Organizational Psychology*, 2, 35-38.
- Campion, M.A., Outtz, J.L., Zedeck, S., Schmidt, F.L., Kehoe, J.F., Murphy, K.R. & Guion, R.M. (2001). The controversy over score banding in personnel selection: Answers to 10 key questions. *Personnel Psychology*, 54, 149-185.

- Chan, D. & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159.
- Christian, M.S., Edwards, B.D. & Bradley, J.C. (2010). Situational Judgment Tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, 83-117.
- Cizek, G. (1999). *Cheating on tests: how to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Clevenger, J., Pereira, G.M., Wiechmann, D., Schmitt, N. & Harvey, V.S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410-417.
- Cunningham, M.R., Wong, D.T. & Barbee, A.P. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid Report. *Journal of Applied Psychology*, 79, 643-658.
- De Corte, W. & Lievens, F. (2003). A practical procedure to estimate the quality and the adverse impact of single-stage selection decisions. *International Journal of Selection and Assessment*, 11, 89-97.
- Dean, M.A., Roth, P.L. & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology*, 93, 685-691.
- Edwards, B.D. & Arthur, W. (2007). An examination of factors contributing to a reduction in subgroup differences on a constructed-response paper-and-pencil test of scholastic achievement. *Journal of Applied Psychology*, 92, 794-801.
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor & Department of Justice (1978). *Uniform guidelines on employee selection procedures*. 29 C.F.R. 1607.
- Farh, J.L. & Werbel, J.D. (1986). Effects of purpose of the appraisal and expectation of validation on self-appraisal leniency. *Journal of Applied Psychology*, 71, 527-529.
- Fetzer, M., Tuzinski, K. & Freeman, M. (2010, April). *3D animation, motion capture, and SJTs: I-O is finally catching up with IT*. Paper presented at the 25th Annual Conference of Industrial and Organizational Psychology, Atlanta, Georgia.
- Foster, D. (2009). Secure, online, high-stakes testing: Science fiction or business reality? *Industrial and Organizational Psychology*, 2, 31-34.
- Foster, D., Maynes, D. & Hunt, B. (2008). Using data forensic methods to detect cheating. In C.L. Wild & R. Ramaswamy (Eds.), *Improving testing: Applying process tools and techniques to assure quality* (pp. 305-322). Mahwah, NJ: Lawrence Erlbaum Associates.
- Greenwald, A.G. (2008). Landy is incorrect: Stereotyping can be moderated by individuating the out-group and by being accountable. *Industrial and Organizational Psychology*, 1, 430-435.
- Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L. & Mahzarin, R.B. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97, 17-41.
- Hanges, P.J. & Ziegert, J.C. (2008). Stereotypes about stereotype research. *Industrial and Organizational Psychology*, 1, 436-438.
- Hattrup, K., Rock, J. & Scalia, C. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *Journal of Applied Psychology*, 82, 656-664.
- Hattrup, K., Schmitt, N. & Landis, R.S. (1992). Equivalence of constructs measured by job-specific and commercially available aptitude tests. *Journal of Applied Psychology*, 77, 298-308.
- Hausknecht, J.P., Day, D.V. & Thomas, S.C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639-683.

- Hogan, R. & Hogan, J. (2002). Hogan Personality Inventory. Swets Test Publishers.
- 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, 152-194.
- Iles, P.A. & Mabey, C. (1993). Managerial career development techniques: Effectiveness, acceptability, and availability. *British Journal of Management*, 4, 103-118.
- International Test Commission (2006). International guidelines on computer-based and Internet delivered testing. *International Journal of Testing*, 6, 143-172.
- James, L.R., McIntyre, M.D., Glisson, C.A., Green, P.D., Patton, T.W., LeBreton, J.M., Frost, B.C., Russel, S.M., Sablinski, C.J., Mitchell, T.R. & Williams, L.J. (2005). A conditional reasoning measure for aggression. *Organizational Research Methods*, 8, 69-99.
- Jodoin, M.G., Zenisky, A. & Hambleton, R.K. (2006). Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education*, 19, 203-220.
- Kanning, U.P., Grewe, K., Hollenberg, S. & Hadouch, M. (2006). From the subject's point of view: Reactions to different types of situational judgment items. *European Journal of Psychological Assessment*, 11, 184-193.
- Kehoe, J. (2000). *Managing Selection in Changing Organizations: Human Resource Strategies*. San Francisco, CA.: Jossey-Bass.
- Landy, F.J. (2008a). Stereotypes, bias, and personnel decisions: Strange and stanger. *Industrial and Organizational Psychology*, 1, 379-392.
- Landy, F.J. (2008b). Stereotyping, implicit association theory, and personnel decisions: I guess we will just have to disagree. *Industrial and Organizational Psychology*, 1, 444-453.
- LeBreton, J.M., Barksdale, C.D., Robin, J. & James, L.R. (2007). Measurement issues associated with conditional reasoning tests: indirect measurement and test faking. *Journal of Applied Psychology*, 92, 1-16.
- Lerner, J.S. & Tetlock, P.E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125, 255-275.
- Lievens, F. (2007). Employer branding in the Belgian army: The importance of instrumental and symbolic beliefs for potential applicants, actual applicants, and military employees. *Human Resource Management*, 46, 51-69.
- Lievens, F., De Corte, W. & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, 93, 268-279.
- Lievens, F., Harris, M.M., Van Keer, E. & Bisqueret, C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology*, 88, 476-498.
- Lievens, F., Peeters, H. & Schollaert, E. (2007). Situational judgment tests: A review of recent research. *Personnel Review*, 37, 426-441.
- Lievens, F. & Sackett, P.R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, 91, 1181-1188.
- Lievens, F. & Schollaert, E. (2008). Naar een nieuwe generatie assessment: Een open boek over situationele tests. Barneveld, Nederland: Uitgeverij Nelissen.
- McBride, J.R. (1998). Innovations in computer-based ability testing: promise, problems, and perils. In M.D. Hakel (Ed.), *Beyond multiple choice. Evaluating alternatives to traditional testing for selection* (pp. 23-39). Mahwah, NJ: Lawrence Erlbaum Associates.

- McConnell, A.R. & Leibold, J.M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental and Social Psychology, 37*, 435-442.
- McDaniel, M.A., Hartman, N.S., Whetzel, D.L. & Grubb, W.L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91.
- McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Campion, M.A. & Braverman, E.P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730-740.
- Meltzer, P.H. (1995). Videotest voor communicatieve vaardigheden. In F.J.R.C. Dochy & T.R. de Rijke (Eds.), *Assessment centers: Nieuwe toepassingen in opleiding, onderwijs en HRM* (p. 109-123). Utrecht, Nederland: Lemma.
- Motowidlo, S.J., Hooper, A.C. & Jackson, H.L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology, 91*, 749-761.
- Nosek, B.A., Greenwald, A.G. & Banaji, M.R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J.A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265-292). New York: Psychology Press.
- Nye, C.D., Do, B.R., Drasgow, F. & Fine, S. (2008). Two-step testing in employee selection: Is score inflation a problem? *International Journal of Selection and Assessment, 16*, 112-120.
- Ones, D.S. & Viswesvaran, C. (1998). Gender, age, and race differences on overt integrity tests: Results across four large-scale job applicant data sets. *Journal of Applied Psychology, 83*, 35-42.
- Ones, D.S., Viswesvaran, C. & Schmidt, F. (1993). Comprehensive meta-analysis of integrity validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679-703.
- Ostrom, J.K., Born, M.P., Serlie, A.W. & Van der Molen, H.T. (2010). Webcam testing: Validation of an innovative open-ended multimedia test. *European Journal of Work and Organizational Psychology, 19*, 532-550.
- Ployhart, R.E. & Holtz, B.C. (2008). The diversity-validity dilemma: Strategies for reducing race/ethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*, 153-172.
- Ployhart, R.E., Weekley, J.A. & Baughman, K. (2006). The structure and function of human capital emergence: A multilevel examination of the attraction-selection-attrition model. *Academy of Management Journal, 49*, 661-677.
- Pulakos, E.D., Arad, S., Donovan, M.A. & Plamondon, K.E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 85*, 612-624.
- Pulakos, E.D., Schmitt, N. & Chan, D. (1996). Models of job performance ratings: An examination of rater race, ratee gender, and rater level effects. *Human Performance, 9*, 103-119.
- Pulakos, E.D., Schmitt, N., Dorsey, D.W., Arad, S., Hedge, J.W. & Borman, W.C. (2002). Predicting adaptive performance: Further tests of a model of adaptability. *Human Performance, 15*, 299-323.
- Reynolds, D.H. & Weiner, J.A. (2009). *Online recruiting and selection: Innovations in talent acquisition*. Malden, MA: Wiley-Blackwell.

- Richman-Hirsh, W.L., Olson-Buchanan, J.B. & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, 85, 880-887.
- Roth, P.L. & Bobko, P. (2000). College grade point average as a personnel selection device: Ethnic group differences and potential adverse impact. *Journal of Applied Psychology*, 85, 399-406.
- Roth, P.L., Bobko, P., McFarland, L. & Buster, M. (2008). Work sample tests in personnel selection: A meta-analysis of black-white differences in overall and exercise scores. *Personnel Psychology*, 61, 637-662.
- Rotundo, M. & Sackett, P.R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, 87, 66-80.
- Rudman, L.A. (2008). The validity of the implicit association test is a scientific certainty. *Industrial and Organizational Psychology*, 1, 426-429.
- Ryan, A.M. & Tippins, N.T. (2004). Attracting and selecting: What psychological research tells us. *Human Resource Management*, 43, 305-318.
- Rynes, S.L. & Barber, A.E. (1990). Applicant attraction strategies: An organizational perspective. *Academy of Management Review*, 15, 286-310.
- Sackett, P.R. & Ellingson, J.E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology*, 50, 707-721.
- 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, 302-318.
- 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, 262-274.
- Schmit, M.J., Ryan, A.M., Stierwalt, S.L. & Powell, A.B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, 80, 607-620.
- Schmitt, N. & Mills, A.E. (2001). Traditional tests and job simulations: Minority and majority performance and test validities. *Journal of Applied Psychology*, 86, 451-458.
- Sommers, S.R. (2006). On racial diversity and group decision making: Identifying multiple effects on racial composition on jury deliberations. *Journal of Personality and Social Psychology*, 90, 597-612.
- Stanton, J.M. (1999). Validity and related issues in web-based hiring. *The Industrial-Organizational Psychologist*, 36, 69-77.
- Stewart, G.L. & Nandkeolyar, A.K. (2006). Adaptation and intraindividual variation in sales outcomes: Exploring the interactive effects of personality and environmental opportunity. *Personnel Psychology*, 59, 307-332.
- Tam, A.P., Murphy, K.R. & Lyall, J.T. (2004). Can changes in differential dropout rates reduce adverse impact? A computer simulation study of a multi-wave selection system. *Personnel Psychology*, 57, 905-934.
- Tippins, N.T. (2009a). Internet alternatives to traditional proctored testing: Where are we now? *Industrial and Organizational Psychology*, 2, 2-13.
- Tippins, N.T. (2009b). Where is the unproctored internet testing train headed now? *Industrial and Organizational Psychology*, 2, 69-76.
- Tippins, N.T., Beaty, J., Drasgow, F., Gibson, W.M., Pearlman, K., Segall, D.O. & Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, 59, 198-225.

- Van Iddekinge, C.H., Eidson Jr., C.E., Kudisch, J.D. & Goldblatt, A.M. (2003). A biodata inventory administered via interactive voice response (IVR) technology: Predictive validity, utility, and subgroup differences. *Journal of Business and Psychology*, 18, 145-156.
- Van Scotter, J.R., Motowidlo, S.J. & Cross, T.C. (2000). Effects of task performance and contextual performance on systemic rewards. *Journal of Applied Psychology*, 85, 526-535.
- Vince, J. (1998). *Essential Virtual Reality Fast*. London: Springer.
- Wanek, J.E. (1999). Integrity and honesty testing: What do we know? How do we use it? *International Journal of Selection and Assessment*, 7, 183-195.
- Whetzel, D.L., McDaniel, M.A. & Nguyen, N.T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21, 291-309.
- Ziegert, J.C. & Hanges, P.J. (2005). Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, 90, 553-562.